

ESAC Opinion on the validation study of the epiCS[®] Skin Irritation Test (SIT) based on the EURL ECVAM/OECD Performance Standards for *in vitro* skin irritation testing using Reconstructed human Epidermis (RhE)

*ESAC Opinion No. 2016-05
of 24 June 2016*

2016



This publication is an Opinion by ESAC, the EURL ECVAM Scientific Advisory Committee, for the Joint Research Centre (JRC), the European Commission's science and knowledge service. The Opinion does not necessarily reflect the views of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

Contact information

Name: João BARROSO
Address: Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
E-mail: Joao.BARROSO@ec.europa.eu
Tel.: +39 0332 78 5329

JRC Science Hub

<https://ec.europa.eu/jrc>

JRC103707

EUR 28177 EN

PDF ISBN 978-92-79-63199-3 ISSN 1831-9424 doi:10.2787/64982

Luxembourg: Publications Office of the European Union, 2016

© European Union, 2016

Reproduction is authorised provided the source is acknowledged.

How to cite: EURL ECVAM Scientific Advisory Committee (2016). ESAC Opinion on the validation study of the epiCS[®] Skin Irritation Test (SIT) based on the EURL ECVAM/OECD Performance Standards for *in vitro* skin irritation testing using Reconstructed human Epidermis (RhE). ESAC Opinion No. 2016-05 of 24 June 2016; EUR 28177 EN; doi:10.2787/64982. Available at: <http://publications.jrc.ec.europa.eu/repository/handle/JRC103707>.

All images © European Union 2016



EUROPEAN COMMISSION

DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE

Directorate F - Health, Consumers and Reference Materials

European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)

ESAC OPINION

on the

Validation Study of the epiCS[®] Skin Irritation Test (SIT)

based on the EURL ECVAM/OECD Performance

Standards for *In Vitro* Skin Irritation Testing using

Reconstructed human Epidermis (RhE)

ESAC Opinion No.	2016-05
Relevant ESAC Request No.	2014-01
Date of Opinion	24/06/2016

Table of contents

Abstract	1
ESAC Opinion.....	2
References	4
Annex 1: Composition of ESAC and ESAC Rapporteurs	5
Annex 2: EURL ECVAM Request for ESAC Advice	7
Annex 3: ESAC Rapporteurs Peer Review Consensus Report	27

Abstract

ESAC, the EURL ECVAM Scientific Advisory Committee, advises EURL ECVAM on scientific issues. Its main role is to conduct independent peer review of validation studies of alternative test methods and to assess their scientific validity for a given purpose. The committee reviews the appropriateness of study design and management, the quality of results obtained and the plausibility of the conclusions drawn. ESAC peer reviews are formally initiated with a EURL ECVAM Request for ESAC Advice, which provides the necessary background for the peer-review and establishes its objectives, timelines and the questions to be addressed. The peer review is normally prepared by specialised ESAC Working Groups. These are typically composed of ESAC members and other external experts relevant to the test method under review. These experts may be nominated by ESAC, EURL ECVAM and partner organisations within the International Cooperation on Alternative Test Methods (ICATM). ESAC ultimately decides on the composition of these Working Groups. ESAC's advice to EURL ECVAM is formally provided as 'ESAC Opinions' and 'Working Group Reports' at the end of the peer review. ESAC may also issue Opinions on other scientific issues of relevance to the work and mission of EURL ECVAM but not directly related to a specific alternative test method.

The ESAC Opinion expressed in this report relates to the peer-review of the validation study of the epiCS[®] Skin Irritation Test (SIT) based on the EURL ECVAM/OECD Performance Standards for *in vitro* skin irritation testing using Reconstructed human Epidermis (RhE).



EUROPEAN COMMISSION

DIRECTORATE-GENERAL
JOINT RESEARCH CENTRE

Directorate F - Health, Consumers and Reference Materials

European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Ispra, 24 June 2016

ESAC Opinion

In 2014, the EURL ECVAM Scientific Advisory Committee (ESAC) (Annex 1) was requested to peer review and offer an opinion as to whether the performance of the epiCS[®] RhE *in vitro* Skin Irritation Test (SIT) satisfies the relevant Performance Standards with a view to consideration of its use for routine dermal irritation testing hazard classification for regulatory purposes. Two ESAC Rapporteurs were appointed by ESAC at its 39th meeting to conduct a detailed scientific peer review of the test method and facilitate the ESAC Opinion requested by EURL ECVAM. The resulting ESAC Opinion (2014-01) of 17 November 2014 recommended that supplementary testing should be conducted at two laboratories to remove potential sources of bias from the chemical testing datasets made available to ESAC in 2014.

The validation study was set up and managed by the test developer. The initial ring-trial was conducted in 2011. Further testing was performed in 2013 as the result of suspected proficiency problems at the testing laboratories. Additional testing was conducted and submitted to EURL ECVAM in 2015 following the recommendations of ESAC in its Opinion of November 2014.

Following this new submission, the ESAC Rapporteurs (Annex 1) were asked by EURL ECVAM (Annex 2 - ESAC Request 2014-01, updated 31/03/2016) to take account of the additional data and analyses provided in a 2015 Report from the test developer in response to the ESAC Opinion. The ESAC Rapporteurs delivered a report dated 6 June 2016 setting out their agreed position taking this additional information in to account (Annex 3).

Having considered the 2015 test developer's Report to EURL ECVAM, the ESAC Rapporteurs requested additional analysis in order to derive a dataset that they consider a more appropriate way to judge the performance of the test method. The results are shown in Table 1 below, including Clopper-Pearson exact 95 %-Confidence Intervals calculated by ESAC.

Specific Recommendations

There were issues with the performance of test kits shipped trans-continentially during the first phase of this study. Although the manufacturer attempted to identify and remedy the causes, there may have been some similar problems in the later part of the study. The manufacturer should investigate this, establish whether this was the case for example by confirming whether the quality control test kit batch-release criterion (i.e., measure of barrier function by incubation with Triton X-100) was/is still met when the test kits were/are received or used by the laboratory, and if necessary remedy the causes of the problems.

Readiness for Standardised/Regulatory Use - Recommendation:

On the basis of the best, albeit still imperfect, dataset currently available, the test method satisfies all but one of the Performance Standards target values: the overall WLR value obtained in one laboratory was 89 %. This value is marginally below the Performance Standards target value of 90 %.

Nevertheless, the ESAC believes that the test method has the potential to meet the PS target values provided that the specific recommendations above are addressed.

Subject to these provisions, the ESAC believes the test method is suitable for screening purposes, and merits consideration for regulatory use.

Of note, the ESAC also makes general recommendations on Performance Standards in a separate Opinion (ESAC, 2016), which are relevant to the current OECD Performance Standards on skin irritation (OECD Series on Testing and Assessment No. 220; OECD, 2015).

Table 1: Re-calculated FINAL Results

Validation Study Results based on Rapporteurs' calculations			PS Criteria Met Yes/No
Lab. #1 – ACS	Sensitivity	100 % (66–100 %)	
	Specificity	80 % (44–97 %)	
	Accuracy	89 % (67–99 %)	
Lab. #2 – HCCR	Sensitivity	90 % (56–100 %)	
	Specificity	80 % (44–97 %)	
	Accuracy	85 % (62–97 %)	
Lab. #3 – IIVS	Sensitivity	75 % (35–97 %)	
	Specificity	80 % (44–97 %)	
	Accuracy	78 % (52–94 %)	
For all three laboratories	Sensitivity	89 % (71–98 %)	Yes (PS ≥ 80 %)
	Specificity	80 % (61–92 %)	Yes (PS ≥ 70 %)
	Accuracy	84 % (72–93 %)	Yes (PS ≥ 75 %)
WLR Lab. #1 – ACS		95 % (74–100 %)	Yes (PS ≥ 90 %)
WLR Lab. #2 – HCCR		100 % (83–100 %)	Yes (PS ≥ 90 %)
WLR Lab. #3 – IIVS		89 % (65–99 %)	No (PS ≥ 90 %)
BLR		88 % (64–99 %)	Yes (PS ≥ 80 %)
% of complete run seq. in Lab. #1 – ACS		95 % (75–100 %)	Yes (PS ≥ 85 %)
% of complete run seq. in Lab. #2 – HCCR		100 % (83–100 %)	Yes (PS ≥ 85 %)
% of complete run seq. in Lab. #3 – IIVS		90 % (68–99 %)	Yes (PS ≥ 85 %)
% of complete run seq. over the three labs		95 % (86–99 %)	Yes (PS ≥ 90 %)

References

- ESAC (2016). ESAC Opinion on the use of Performance Standards to evaluate test methods similar to a Validated Reference Method. ESAC Opinion 2016-06 of 24 June 2016.
- OECD (2015). Performance Standards for the assessment of proposed similar or modified *in vitro* Reconstructed human Epidermis (RhE) test methods for skin irritation testing as described in TG 439 (Intended for the developers of new or modified similar test methods). Series on Testing and Assessment No. 220.



Annex 1

COMPOSITION OF ESAC AND ESAC RAPPORTEURS



Composition of ESAC and ESAC Rapporteurs

EURL ECVAM Scientific Advisory Committee (ESAC)

- Dr. Neil CARMICHAEL (ESAC Chair)
- Prof. Jürgen BORLAK
- Dr. Harvey CLEWELL
- Prof. Lucio G. COSTA
- Dr. Kristina KEJLOVÁ
- Prof. David John KIRKLAND
- Prof. Annette KOPP-SCHNEIDER
- Dr. Renate KRÄTKE
- Prof. Claus-Michael LEHR
- Dr. José Maria NAVAS
- Prof. Aldert PIERSMA
- Dr. Jonathan RICHMOND
- Dr. Erwin L. ROGGEN
- Dr. Dorothea SESARDIC

ESAC Rapporteurs

- Dr. Jon RICHMOND (ESAC Member, Lead drafter)
- Dr. Renate KRÄTKE (ESAC Member)

EURL ECVAM (Secretariat)

- Dr. João BARROSO (ESAC Coordinator)
- Prof. Maurice WHELAN (Head of Unit)



Annex 2

EURL ECVAM REQUEST FOR ESAC ADVICE



ESAC Request 2014-01 (updated 31/03/2016)

EURL ECVAM Scientific Advisory Committee (ESAC)

EURL ECVAM REQUEST FOR ESAC ADVICE

on the

Validation study of the epiCS® Skin Irritation Test (SIT) based on the EURL ECVAM/OECD Performance Standards for *in vitro* skin irritation testing using Reconstructed human Epidermis (RhE)

Title page information	
Abbreviated title of ESAC request	PS-based validation of epiCS® SIT
ESAC REQUEST No.	2014-01
Template used for preparing request	EP 3.02
Date of finalising request	20/02/2014 Updated request: 17/03/2016
Date of submitting request to ESAC	21/02/2014 Updated request: 17/03/2016
Request discussed through	ESAC39 Updated request: Written procedure previous to ESAC 42 (approved 31/03/2016)
Opinion expected at (date)	Written procedure following ESAC 39 Updated request: ESAC 42 (June 2016)
File name of this request	ER2014-01_ESAC REQUEST_epiCS_2014_CLEAN-final.doc Updated request: ER2014- 01_ESAC_UPDATED_REQUEST_epiCS_SIT_2016.doc

TABLE OF CONTENTS

1.	TYPE OF REQUEST	10
2.	TITLE OF STUDY OR PROJECT FOR WHICH SCIENTIFIC ADVICE OF THE ESAC IS REQUESTED.....	10
3.	BRIEF DESCRIPTION OF THE STUDY OR PROJECT	10
4.	OBJECTIVES, QUESTIONS, TIMELINES.....	17
4.1	OBJECTIVE	17
4.2	QUESTION(S) TO BE ADDRESSED	18
4.3	TIMELINES	19
5.	EURL ECVAM PROPOSALS ON HOW TO ADDRESS THE REQUEST WITHIN ESAC.....	19
5.1	EURL ECVAM PROPOSAL REGARDING REQUEST-RELATED STRUCTURES REQUIRED	19
5.2	DELIVERABLES AS PROPOSED BY EURL ECVAM	20
6.	LIST OF DOCUMENTS TO BE MADE AVAILABLE TO THE ESAC	20
7.	TERMS OF REFERENCE OF THE ESAC RAPPORTEURS	24
7.1	ESTABLISHMENT OF THE ESAC RAPPORTEURS.....	24
7.2	TITLE OF THE ESAC RAPPORTEURS.....	24
7.3	MANDATE OF THE ESAC RAPPORTEURS	24
7.4	DELIVERABLES OF THE ESAC RAPPORTEURS	24
7.5	PROPOSED TIMELINES OF THE ESAC RAPPORTEURS	24
7.6	QUESTIONS WHICH SHOULD BE ADDRESSED BY THE ESAC RAPPORTEURS.....	25
APPENDIX 1	REPORTING TEMPLATE	25

1. TYPE OF REQUEST

Request Type	Identify request ("YES")
R1 ESAC Peer Review of a Prevalidation Study or Validation Study	YES, external validation study (i.e. not coordinated by EURL ECVAM)
<i>If R1)applies please specify further:</i>	
► Prevalidation Study	NO
► Prospective Validation Study	NO
► Retrospective Validation Study	NO
► Validation Study based on Performance Standards	YES The validation study is based on the Performance Standards (PS) for <i>in vitro</i> Skin Irritation testing using Reconstructed human Epidermis (RhE). The PS are outlined in the EURL ECVAM Performance Standards (EURL ECVAM 2009) and in the OECD Series on Testing and Assessment no. 220 (OECD 2015a). The first version of the OECD PS was included as an Annex in Test Guideline 439, which was first adopted in 2010 (OECD 2010). The validation study was performed by the test method developer.
R2 Scientific Advice on a test method submitted to EURL ECVAM for validation (e.g. the test method's biological relevance etc.)	NO
R3 Other Scientific Advice (e.g. on test methods, their use; on technical issues such as cell culturing, stem cells, definition of performance standards etc.)	NO

2. TITLE OF STUDY OR PROJECT FOR WHICH SCIENTIFIC ADVICE OF THE ESAC IS REQUESTED

epiCS® Skin Irritation Study

3. BRIEF DESCRIPTION OF THE STUDY OR PROJECT

Important note:

During the study phase the test submitter changed the name of the test method from EST1000 SIT to epiCS® SIT (SIT=Skin Irritation Test). The submitter has assured EURL ECVAM that the name change constitutes solely a change of trade name and that the test system (RhE) and associated SOP (way of conducting the test) have remained identical, i.e. not undergone any modification.

Summary

The epiCS® (formerly known as EST1000) is a 3-dimensional *Reconstructed human Epidermis* (RhE) model consisting of normal primary human epidermal keratinocytes from one donor. The production process using defined media conditions leads to the differentiation of the cells and the formation of a multilayered epidermis. It consists of basal, spinous and granular layers and a multilayered stratum corneum highly comparable to native human skin. The epiCS® model was already validated for **skin corrosion testing** by EURL ECVAM in 2009 (ESAC 2009) and is accepted by the OECD for the classification of chemicals causing skin corrosion (OECD TG 431) (OECD 2015c). The current ESAC request concerns the scientific review of a submission of a validation study that was performed on the basis of **Performance Standards (PS) for *in vitro* skin irritation testing using RhE**. The relevant PS have been defined by EURL ECVAM (2009). They were also published in the OECD Series on Testing and Assessment No. 220 in 2015 (OECD 2015a), which accompany OECD TG 439 (OECD 2015b); the first version of the OECD PS was included as an Annex in Test Guideline 439, which was first adopted in 2010 (OECD 2010).

The PS-based validation study to be reviewed was conducted by the test method developer (CellSystems, Germany) in between 2011 and 2015 (see below, 'History of submissions') and the final study and associated data were submitted to EURL ECVAM in Q4 2015. Like previously validated *in vitro* RhE-based skin irritation methods, the epiCS® Skin Irritation Test (SIT) is proposed to be capable of discriminating between not-classified (GHS No Category) and classified (GHS Categories 1 or 2) chemicals. It is not able to distinguish between irritant (GHS Category 2) and corrosive (GHS Category 1) chemicals nor does it generate data on the optional GHS Category 3 ('mild irritants'). As GHS Category 3 is not implemented in the EU and other global regions, the method may serve in these regions as a full replacement to the skin irritation part of the traditional rabbit Draize skin test (OECD TG 404) (OECD 2015d), thus reducing the use of laboratory animals.

The epiCS® SIT was first submitted to EURL ECVAM in 2009/2010 for an assessment of its compliance with the Essential Test Method components of the PS for RhE-based *in vitro* skin irritation test methods (EURL ECVAM 2009; OECD 2010; OECD 2015a). EURL ECVAM and ESAC confirmed that the epiCS® SIT qualified for a PS-based ("catch-up") validation study. The experimental ring trials (submitted in 2011 and, after retesting, in 2013 and 2015) had then the objective of assessing whether the epiCS® SIT meets the performance criteria in terms of reproducibility within and between laboratories and in terms of predictive capacity (sensitivity, specificity and accuracy of predictions).

History of submissions

a) In 2009, CellSystems submitted the EURL ECVAM Test Method "**Proposal Evaluation Form**", providing a short outline and description of the EST-1000 SIT test method. The proposal Evaluation Form served, at the time, as a **presubmission step** at EURL ECVAM. Having evaluated the proposed EST-1000 method, EURL ECVAM informed the submitter that the test method may qualify for a Performance-Standards based validation study depending however on a complete evaluation & conclusion of sufficient method similarity between the EST-1000 assay and the validated reference methods. The submitter was invited to submit a more comprehensive test method description in the EURL ECVAM "Test Submission Template" providing also historical data on test method development, batch stability, etc.

b) In 2010 a detailed test method description was submitted in agreement with the provisions of EURL ECVAM's PS on skin irritation testing. This test method description addressed the essential test method components outlined in the EURL ECVAM Performance Standards (PS) for *in vitro* skin irritation test methods based on Reconstructed human Epidermis (RhE) (EURL ECVAM 2009)¹.

¹ An abridged version of these PS is reproduced as Annex 2 of OECD TG 439. The PS of TG439 are identical to those of ECVAM but do not provide some of the detailed guidance contained in the ECVAM PS.

According to EURL ECVAM's approach for the validation of putative similar methods, confirmation of **sufficient similarity** (based on analysis of the essential test method components) of a putative **similar method** ('me-too method') with respect to the **validated reference method(s)**, is a prerequisite for entering EURL ECVAM validation based on PS ("catch-up validation"). The assessment performed by EURL ECVAM in 2009 indicated **full compliance with these criteria** and ESAC (in Q1 2011) agreed with this assessment, confirming sufficient similarity to validate the test method using the Reference Chemicals and Target Reproducibility and Accuracy Values defined in the Performance Standards (EURL ECVAM 2009; OECD 2010; OECD 2015a). The TST also contained the current SOP and in-house testing data from the test method developer (optimisation & testing) providing preliminary information on the predictive capacity and reproducibility of the EST-1000 method. EURL ECVAM informed the submitter that the test method would qualify for PS-based validation and invited the submitter to submit a revised TST, once the PS-based ring trial employing three laboratories had been completed.

c) In 2011, the Test Submitter completed a PS-based validation study based on a ring trial with three laboratories and submitted the documents to EURL ECVAM for assessment. The PS-based study addressed modules 1 to 6 (EURL ECVAM's modular approach). On the basis of the submitted results, EURL ECVAM concluded that the test method was not ready to progress to ESAC peer review as the target values for **Within Laboratory Reproducibility** (as defined in the PS) had not been met by any of the three participating laboratories: the concordance values observed were 74%, 84% and 84% (target value in PS is 90%). In addition, the **Between Laboratory Reproducibility** was slightly below the target value defined in the PS (78% compared to 80%, respectively). In contrast the **predictive capacity** target value was met (overall accuracy of 85%; PS stipulate accuracy $\geq 75\%$), and thus exceeded the performance of the validated reference method (VRM) (80 %).

In ensuing communications EURL ECVAM and the submitter agreed that the test method appeared capable of discriminating not-classified from classified (irritants/corrosives) chemicals (predictive capacity met) but that there were issues with the reproducibility of the method (WLR and BLR) which might be due to *inter alia* transfer problems (e.g. shipment, lack of training) or intrinsic batch variability.

Since the predictive capacity target value was met and since the BLR was very close to the threshold (2% difference), it was agreed that batch variability was rather unlikely (in this case PC and BLR would be expected to be affected to the same extent as WLR, which was not the case) and that probably issues of protocol transfer / training had caused the comparably low WLR values (especially in the naïve laboratory (ACS)), i.e. that insufficient attention had been given to ensure transferability of the method, in particular to the naïve laboratory. This assumption was further supported by the fact that no transfer phase had been conducted before the ring trial. Since the test developer assumed that both IIVS and Harlan were experienced users of RhE test methods, proficiency testing only with the negative and positive control had been conducted. As a consequence also the naïve laboratory (ACS) had not received transfer training and had not checked with chemicals other than positive and negative controls proficiency of conducting the method. It is therefore plausible that transfer issues had caused reproducibility problems. A small scale transfer phase (e.g. using the controls and a few test chemicals (e.g. $n=4$) or using the proficiency chemicals² (TG 439) could have served to ensure successful transfer, or indicated already at this step issues.

Therefore, EURL ECVAM recommended the following re-testing strategy: As the greatest variability had been observed at the naïve laboratory, all 20 reference chemicals should be retested in three runs. The two other laboratories should at least retest the six reference chemicals that had given discordant results in three runs. Additionally, one of these labs had to re-test the chemical that

² Proficiency chemicals are used to ensure that a laboratory that is implementing a test method is indeed proficient in conducting the method. In any end user laboratory (e.g. CROs) proficiency should be assessed before routine testing commences and proficiency should moreover be assessed at regular intervals and whenever there are changes to laboratory personnel, equipment or testing conditions (e.g. new air condition, air flow etc.).

created an incomplete run sequence (1-decanol) in three runs. The data obtained after this retesting could then be compiled with the original data to arrive at a complete data matrix. It was communicated to the Submitter that this approach was only a recommendation and that the responsibility for choosing the appropriate approach to validate the test method was with the Submitter.

d) In November 2013 a revised full dossier was submitted to EURL ECVAM containing a data matrix composed as recommended by EURL ECVAM (described above). Briefly, the results obtained were:

- **Occurrence of non-qualified tests/runs:** A run is composed of the testing of one or several test chemicals in triplicate (tests) plus the testing of positive and negative controls. If the positive or negative controls are outside of the accepted ranges (ViabilityPC \leq 20%; $1.0 \leq$ ODNC \leq 2.8) the run is considered not qualified and, consequently, all tests included within that run are also considered not qualified. The mean of triplicate values of each test is also calculated and if the SD > 18 %, the test is considered not qualified. Twenty out of the 198³ tests conducted (10 %) were non-qualified (see statistical report, table on page 14). In comparison to the original study of 2011 (18.4 % non-qualified runs), the occurrence of non-qualified tests appeared to decrease. The study acceptance criteria as stipulated in the PS (Section 3.3.1, paragraph 21) were met.
- **Complete run sequences:** two of the three laboratories had complete run sequences for all 20 chemicals following re-testing (100 %). The naïve laboratory (ACS) had complete run sequences for 19 out of 20 chemicals (95 %), thus meeting the acceptance criterion of \geq 85 %. An incomplete run sequence was obtained for 1-methyl-3-phenylpiperazine since, in contrast to the last study (in 2009), ACS was unable to generate three qualified runs after a maximum of 5 independent runs (up to two re-tests). BLR could therefore be assessed only on the basis of 19 chemicals (PS outline that BLR should only be conducted for chemicals with valid run sequences in all laboratories). In total, 98.3 % of the run sequences were complete, thus meeting the acceptance criterion of \geq 90 %.
- **Within-laboratory reproducibility (WLR):** 95 % in ACS, 100 % in HCCR and 95% in IIVS (based on concordance of predictions of runs)
- **Between-laboratory reproducibility (BLR):** 19 of the 20 chemicals were predicted concordantly by the three laboratories (95% concordance). For one chemical valid predictions were available only in 2 of 3 labs. This substance, in agreement with the PS, was not considered for BLR calculation. Of the 19 concordant predictions 17 were correct with respect to the in vivo reference data, while 2 were FPs.
- **False negative results** were obtained for 1-methyl-3-phenylpiperazine at HCCR only, thus meeting the acceptance criteria for false negative results described in the PS. ACS produce 1 qualified true positive result and 3 non-qualified false negative results for 1-methyl-3-phenyl-1-piperazine, with the mean of the four runs also being false negative.
- **False positive results** were obtained for allyl-phenoxy-acetate and cinnamaldehyde in all participating laboratories.
- The **sensitivity of ca 97 %** (100 % at ACS and IIVS; 90 % at HCCR) was above the sensitivity of the VRM (ca 86 %) and thus met the acceptance criterion of \geq 80 %.
- The **specificity of 80 %** (80 % in all participating laboratories) was above the VRM (ca 77 %) and thus met the acceptance criterion of \geq 70 %.
- The **overall accuracy of ca 88 %** (89.5 % at ACS, 85 % at HCCR, 90 % at IIVS) was above the VRM (ca 80 %) and thus met the acceptance criterion of \geq 75 %.

³ This is the number of tests conducted per chemical during the study. The maximally possible number of tests is 5 per chemical (3 tests plus 2 retests if required). Therefore, the maximal number of tests per study is 5 x 20 (reference chemicals) x 3 Laboratories = 300 tests.

- Some minor deviations from PS were observed concerning the time points & batches used for generating the data on reference chemicals.

e) In October 2014, at its 40th plenary meeting, the ESAC expressed concerns on the way the second validation study of 2013 had been conducted. In particular, the ESAC questioned the appropriateness of the study design. These concerns were related to partial re-testing, in 2013, of the reference chemicals in two out of three laboratories (HCCR and IIVS), while the leading laboratory (ACS) re-tested the entire set of 20 reference chemicals. At HCCR and IIVS, only 7 and 6 of the reference chemicals, respectively, had been re-tested in 2013. Hence, for these two laboratories, the data generated in 2011 and 2013 were grouped across the 20 chemicals. The ESAC deemed that grouping data introduced a bias in the data set generated from these two laboratories, since only non-concordant data points had been re-tested. The ESAC therefore recommended that the remaining 13 and 14 chemicals be re-tested respectively by HCCR and IIVS in order to complete the 2013 validation study and obtain an unbiased complete dataset. These remaining chemicals were re-tested in 2015.

f) In October 2015, Cell Systems submitted to EURL ECVAM a full validation study complying with ESAC's recommendations, i.e. containing new data for the 13 and 14 chemicals that had not been re-tested respectively by HCCR and IIVS in the validation trial conducted in 2013. These data were submitted to EURL ECVAM together with a report that the submitter prepared on this re-testing.

Different phases of the PS-based validation of the epiCS® SIT

ACS		HCCR		IIVS	
		2011			
20 chemicals		20 chemicals		20 chemicals	
ACS		HCCR		IIVS	
		2013 + 2015			
2013		2013		2013	
20 chemicals re-tested		7 chemicals re-tested		6 chemicals re-tested	
		2015		2015	
		13 chemicals re-tested		14 chemicals re-tested	

Briefly, the results obtained with the pooled dataset from the studies of 2013 and 2015 are:

- **Occurrence of non-qualified tests/runs:** A run is composed of the testing of one or several test chemicals in triplicate (tests) plus the testing of positive and negative controls. If the positive or negative controls are outside of the accepted ranges ($\text{Viability}_{\text{PC}} \leq 20\%$; $1.0 \leq \text{OD}_{\text{NC}} \leq 2.8$) the run is considered not qualified and, consequently, all tests included within that run are also considered not qualified. The mean of triplicate values of each test is also calculated and if the SD > 18 %, the test is considered not qualified. Thirty four out of the 212 tests conducted (16 %) were non-qualified. Of these 34 non-qualified tests, 16 occurred due to failure of the positive control rather than SD > 18 %. The occurrence of non-qualified tests is

similar to what was observed in the original study of 2011 (18 % non-qualified tests). The study acceptance criteria as stipulated in the PS (Section 3.3.1, paragraph 21) were met.

- **Complete run sequences:** two of the three laboratories have complete run sequences for all 20 chemicals following re-testing (100 %). The naïve laboratory (ACS) has complete run sequences for 19 out of 20 chemicals (95 %), thus meeting the acceptance criterion of ≥ 85 %. An incomplete run sequence was obtained for 1-methyl-3-phenyl-1-piperazine since ACS was unable to generate three qualified runs after a maximum of 5 independent runs (up to two re-tests). In total, 98.3 % of the run sequences are complete, thus meeting the acceptance criterion of ≥ 90 %.
- The **WLR of 94.7 % at ACS, 100 % at HCCR and 90 % at IIVS** meets the acceptance criterion of ≥ 90 %.
- The **BLR of 84.2 %** meets the acceptance criterion of ≥ 80 %. Since the reference chemical 1-methyl-3-phenyl-1-piperazine did not have complete run sequences in all laboratories, it was not considered for the calculation of BLR, according to the provisions of the PS (OECD 2015a). It should however be noted that a qualified false negative result and a qualified true positive result were obtained for this chemical at HCCR and IIVS, respectively, and therefore, the chemical did not show reproducible results between laboratories independently of the results obtained by ACS (where an incomplete run sequence was obtained). If this chemical was considered in the calculation of BLR, the value obtained would be 80 %.
- The **sensitivity of 86.2 %** (100 % at ACS, 90 % at HCCR, 70 % at IIVS) meets the acceptance criterion of ≥ 80 %.
- **False negative results** (based on the mean viability of the three qualified runs) were obtained for 1-methyl-3-phenyl-1-piperazine at HCCR, and for 1-bromohexane, 1-decanol and di-n-propyl disulphide at IIVS, thus meeting the acceptance criteria for false negative results described in the PS. ACS produce 1 qualified true positive result and 3 non-qualified false negative results for 1-methyl-3-phenyl-1-piperazine, with the mean of the four runs also being false negative.
- The **specificity of 80 %** (80 % in all participating laboratories) meets the acceptance criterion of ≥ 70 %.
- **False positive results** (based on the mean viability of the three qualified runs) were obtained for allyl-phenoxy-acetate and cinnamaldehyde in all three participating laboratories.
- The **overall accuracy of 83.1 %** (89.5 % at ACS, 85 % at HCCR, 75 % at IIVS) meets the acceptance criterion of ≥ 75 %.
- A significant deviation from the PS was observed concerning the maximum number of re-tests allowed. The PS specify that *"to complement missing data, a maximum of two additional runs... ("re-testing")... may be conducted for each Reference Chemical in each laboratory. Non-qualified tests should be documented and reported"* (OECD 2015a). This means that each laboratory should not test the same reference chemical more than five times. However, 1-decanol and potassium hydroxide (5% aq) were tested 7 and 6 times, respectively, at IIVS because two extra runs not reported in the EURL ECVAM reporting template were not qualified: one, the second overall trial, due to failure of the positive control and the other, the fifth overall trial, due to contamination and failure of the positive control. It should also be noted that a true positive result (with SD < 18 %) was obtained for 1-decanol in trial 2 and, if this result was accepted, this chemical would have non-concordant results within IIVS, thus lowering the laboratory's WLR to 85 %, which is lower than the acceptance criterion of ≥ 90 %.

With these new data and information, the submitter is requesting that ESAC reconsiders/revise its opinion of 2014.

References

- Botham, P.A., Earl, L.K., Fentem, J.H., Roquet, R., Van DE Sandt, J.J.M., 1998. Alternative methods for skin irritation testing: the current status, ECVAM Skin Irritation Task Force Report 1. ATLA 26, 195-211.
- Cotovio, J., Grandidier, M-H., Portes, P., Roguet, R., Rubinstenn, G. 2005. The In Vitro Acute Skin Irritation of Chemicals: Optimisation of the EPIKIN Prediction Model within the Framework of the ECVAM Validation Process. ATLA 33, 329–349.
- Draize J.H., Woodand G. & Calvery H.O. 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. Journal of Pharmacology and Experimental Therapeutics. 82, pp 337-390.
- ECETOC, 1995. Skin Irritation and Corrosion: Reference Chemicals Data Bank ECETOC Technical Report No. 66. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.
- EURL ECVAM, 2009. Performance Standards For In-Vitro Skin Irritation Test Methods Based On Reconstructed Human Epidermis.
- Eskes, C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I. and Zuang, V. 2007. The ECVAM international validation study on in vitro tests for acute skin irritation: selection of test chemicals. ATLA 35, 603-619
- Faller, C., Bracher, N., Dami, N., Roguet, 2002. Predictive ability of reconstructed human epidermis equivalents for assessment of skin irritation of cosmetics. Toxicology in Vitro 16 (5), 557-572.
- Faller, C., Bracher, M., 2002. Reconstructed skin kits: Reproducibility of cutaneous irritating testing. Skin Pharmacol Appl Skin Physiol 15 (suppl 1), 74-91.
- Fentem, J.H., Briggs, D., Chesné, C., Elliott, G.R., Harbell, J.W., Heylings, J.R., Portes, P., Roguet, R., Van De Sandt, J.J.M., Botham, P.A., 2001 A prevalidation study on in vitro tests for acute skin irritation: results and evaluation by the Management Team. Toxicology in Vitro 15, 57-93.
- Griesinger, C., Barroso, J., Zuang, V., Cole, T., Genschow, E. and Liebsch M. 2009. Explanatory background document to the OECD draft test guideline on in vitro skin irritation.
- Heylings, J.R., Diot,., Esdaile, D.J., Fasano, W.J., Manning, L.A., Owen H.M. 2003. A prevalidation study on the in vitro skin irritation function test (SIFT) for prediction of acute skin irritation in vivo: Results and evaluation of ECVAM Phase III. Toxicology in Vitro 17, 123–138.
- Hoffmann J., Heisler E., Karpinski S., Losse J., Thomas D., Siefken W., Ahr HJ., Vohr HW., Fuchs HW 2005. Epidermal-skin-test 1000 (EST-1000)--a new reconstructed epidermis for in vitro skin corrosivity testing. Toxicol In Vitro. 2005 Oct;19(7):925-9. Epub 2005 Aug 2
- Kandárová H., Liebsch M., Genschow E., Ingrid Gerner, Traue D., Slawik B., Spielmann H. 2004: Optimisation of the EpiDerm Test Protocol for the upcoming ECVAM Validation Study on In Vitro Skin Irritation Tests. ALTEX 21, 3/04, 107-114.
- Kandárová, H., Liebsch, M., Gerner, I., Schmidt, E., Genschow, E., Traue, D., Spielmann, H., 2005. EpiDerm Skin Irritation Test Protocol – Assessment of the performance of the optimised test. ATLA 33 (4), p. 351-367.
- Kandárová H., Hayden P., Spiller E., Klausner M., Kubilus J. and Sheasgreen J. 2007: In Vitro Skin Irritation Test: Increasing the Sensitivity of the EpiDerm Skin Irritation Protocol Evaluated in the ECVAM Skin Irritation Validation study. In: Toxicology Letters, Abstracts of the 44th Congress of the European Societies of Toxicology Volume 172, Supplement 1, 7 October 2007, Page 81.
- Mossman, T., 1983. Rapid colorimetric assay for cell growth and survival application to proliferation and cytotoxicity assays. Journal of Immunological Methods 65, 55-63.
- OECD, 2005. OECD Guidance Document No. 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. Organisation for Economic Cooperation and Development, Paris.
- OECD, 2010. OECD Guideline for the Testing of Chemicals, No. 439: *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method (outdated and replaced by newer versions). Organisation for Economic Cooperation and Development, Paris.

- OECD, 2015a. OECD Series on Testing and Assessment, No. 220: Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Reconstructed Human Epidermis (RhE) Test Methods for Skin Irritation Testing as Described in TG 439. Organisation for Economic Cooperation and Development, Paris.
- OECD, 2015b. OECD Guideline for the Testing of Chemicals, No. 439: *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method. Organisation for Economic Cooperation and Development, Paris.
- OECD, 2015c. OECD Guideline for Testing of Chemicals, No. 431: In Vitro Skin Corrosion: Human Skin Model Test. Organisation for Economic Cooperation and Development, Paris, p.8.
- OECD, 2015d. OECD Guideline for the Testing of Chemicals, No. 404: Acute Dermal Irritation/Corrosion. Organisation for Economic Cooperation and Development, Paris.
- Portes, P., Grandidier, M.-H., Cohen, C., Roguet, R., 2002 Refinement of the Episkin® protocol for the assessment of acute skin irritation of chemicals: follow-up to the ECVAM prevalidation study. *Toxicology in Vitro* 16, 765-770.
- Robinson, M.K., Cohen, C., de Brugerolle de Fraissinette, A., Ponec, M., Whittle, E., Fentem, J.H., 2002 Non-animal testing strategies for assessment of the skin corrosion and skin irritation potential of ingredients and finished products. *Food and Chemical Toxicology* 40, 573-592.
- Van de Sandt, J., Roguet, R., Cohen, C., Esdaile, D., Ponec, M., Corsini, E., Barker, C., Fusenig, N., Liebsch, M., Benford, D., de Brugerolle de Fraissinette, A., Fartasch, M., 1999 The Use of Human Keratinocytes and Human Skin Models for Predicting Skin Irritation. The Report and Recommendations of ECVAM Workshop 38. *ATLA* 27, 723-743.
- Zuang, V., Balls, M., Botham, P.A., Coquette, A., Corsini, E., Curren, R.D., Elliott, G.R., Fentem, J.H., Heylings, J.R., Liebsch, M., Medina, J., Roguet, R., Van De Sandt, J.J.M., Wiemann, C., Worth, A.P., 2002 Follow up to the ECVAM prevalidation study on in vitro tests for acute skin irritation, ECVAM Skin Irritation Task Force Report 2. *ATLA* 30, 109-129.

4. OBJECTIVES, QUESTIONS, TIMELINES

4.1 OBJECTIVE

<p>Objective</p> <p><i>Why does EURL ECVAM require advice on the current issue?</i></p>	<p>The opinion of ESAC should conclude on the quality of the submitted Performance Standards-based validation study on the epiCS® SIT, which addressed the reliability (transferability, within and between laboratory reproducibility) and the predictive capacity of the epiCS® SIT. The opinion should conclude on the equivalence of the performance of epiCS® SIT to that of the validated reference methods as outlined in the Performance Standards. Overall, the ESAC opinion, based on the review of the submitted study dossier, should conclude on the adequacy of the epiCS® SIT for routine dermal irritation testing for regulatory purposes.</p>
--	---

4.2 QUESTION(S) TO BE ADDRESSED

<p>Questions</p> <p><i>What are the questions and issues that should be addressed in view of achieving the objective of the advice?</i></p>	<p>UPDATED REQUEST</p> <p>(1) Background:</p> <p>In 2014, the ESAC was requested to review whether the validation study was conducted appropriately in view of the objective of the study. The study objective was to assess the following parameters in comparison to respective criteria outlined in the Performance Standards:</p> <ul style="list-style-type: none"> (a) the reproducibility of the epiCS® SIT within laboratories (WLR) (b) its transferability to other laboratories (c) its reproducibility between laboratories (BLR) (d) its predictive capacity <p>(2) Questions to ESAC (March 2016):</p> <p>The ESAC is now requested to reconsider its previous scientific review of the epiCS® SIT and to revise the ESAC Rapporteur Report and Opinion of November 2014, if/where appropriate, on the basis of the new data and information submitted to EURL ECVAM in October 2015. In particular, the ESAC is requested to assess the new data and information against the criteria set out in the Performance Standards on <i>in vitro</i> skin irritation testing based on RhE and to conclude:</p> <ul style="list-style-type: none"> (a) whether the re-testing performed in 2015 was adequately conducted, especially in light of the previous ESAC opinion (b) whether the final dataset including the data generated after the re-testing conducted in 2015 is complete and unbiased. This item also includes adequate reporting of the re-testing data (e.g., adequate reporting of non-qualified runs, controls and chemicals) (c) whether the number of re-tests/re-runs is acceptable (d) whether the reproducibility of the epiCS® SIT within each of the participating laboratories (WLR) is acceptable (e) whether the reproducibility of the epiCS® SIT between laboratories (BLR) is acceptable (f) whether the predictive capacity of the epiCS® SIT is acceptable (g) whether the overall performance of the method is sufficient in view of its use for routine dermal irritation testing for regulatory purposes <p>N.B. Similarity of the epiCS® had been confirmed previously and, therefore, the essential test method components relating to the test system do not need to be reviewed.</p>
--	--

4.3 TIMELINES

Timelines concerning this request <i>When does EURL ECVAM require the advice?</i>	Timeline	Indication
	Finalised ESAC Opinion required by:	April 2014 Updated opinion: June 2016
	Request to be presented to ESAC by written procedure (e.g. <u>due to urgency</u>) prior to the next ESAC	YES
	Request to be presented to ESAC at ESAC plenary meeting	ESAC Request discussed at ESAC39 11/12 March 2014. Updated ESAC Request of March 2016 shall not be discussed at an ESAC plenary meeting

5. EURL ECVAM PROPOSALS ON HOW TO ADDRESS THE REQUEST WITHIN ESAC

5.1 EURL ECVAM PROPOSAL REGARDING REQUEST-RELATED STRUCTURES REQUIRED

Specific structures required within ESAC to address the request <i>Does the advice require an ESAC working group, an ESAC rapporteur etc.?</i>	Structure(s) required	Required according to EURL ECVAM? (YES/NO)
	S1 ESAC Rapporteur	YES
	S2 ESAC Working Group	NO
	S3 Invited Experts	NO
	<i>Ad S3: If yes – list names and affiliations of suggested experts to be invited and specify whether these are member of the EEP</i>	Jon Richmond (ESAC member) – lead drafter Renate Krätke (ESAC member)
	If other than above (S1-S3):	

5.2 DELIVERABLES AS PROPOSED BY EURL ECVAM

Deliverables	Title of deliverable other than ESAC opinion	Required? (YES/NO)
<i>What deliverables (other than the ESAC opinion) are required for addressing the request?</i>	D1 ESAC Rapporteur Report and draft opinion	YES (a revision of the report and opinion from 2014)
	D2 ESAC Peer Review Report and draft opinion	NO
	If other than above (D1-D2):	

6. LIST OF DOCUMENTS TO BE MADE AVAILABLE TO THE ESAC

Count	Description of document/document set	Already available? (YES/NO)	File name
FOLDER: 1. DOCUMENT SET: Submission 2013			
1	Standard Operating Procedure for EPIDERMAL SKIN TEST 1000 (EST1000) SKIN IRRITATION TEST, version 1.0 of June 2011	Yes	1a_INVITTOX_Protocol_EST1000_SIT_SOP.pdf
2	Standard Operating Procedure for epiCS® in vitro skin irritation, INVITTOX Protocol, version 4.0 of November 2012	Yes	1b_INVITTOX_Protocol_epiCS_SIT_SOP.pdf
3	List of test chemicals	Yes	2_EST1000_SIT_chemicals_set.xls
4	Study Plan for the catch-up validation study on in vitro skin irritation using epidermal skin test 1000 (EST1000)- Version 3 for the initial study (2011)	Yes	13a_EST1000_SIT_Study_plan_initial.pdf
5	Test Plan: Study on in vitro skin irritation using epiCS; additional testing, version 2.0 (2013)	Yes	13b_EST1000_SIT_Study_plan_completed.pdf
6	OECD Testguideline 439 "In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method", version adopted 07/2010	Yes	15_OECD_TG_439_2010.pdf
7	List of 24 references	Yes	16a_References_List_EST1000_SIT.pdf

8	<u>Scientific publication:</u> Hoffmann J, Heisler E, Karpinski S, Losse J, Thomas D, Siefken W, Ahr HJ, Vohr HW, Fuchs HW (2005). Epidermal-skin-test 1000 (EST-1,000)--a new reconstructed epidermis for in vitro skin corrosivity testing. Toxicol In Vitro. Oct;19(7):925-9	Yes	16b_Hoffmann_EST1000.pdf
9	Statistical Report from BfR for the initial study: "Evaluation of the Catch-up Validation study of the Epidermal Skin Test 1000 (EST1000) for in vitro skin irritation testing", version of 14 November 2011	Yes	17a_Statistical_Report_EST1000_initial.pdf
10	Statistical Report from seh consulting for the completed study: "Evaluation of the Catch-up Validation study of epiCS® (formerly: Epidermal Skin Test 1000 (EST1000)) for in vitro skin irritation testing", version of 5 September 2013	Yes	17b_Statistical_Report_EST1000_completed.pdf
11	Lot Release Certificate Epidermal Skin Test 1000 EST1000 as an example for QA	Yes	17c_Lot_Release_Certificate.pdf
12	Letter ECVAM-Test submitter of 19.12.2012; on agreed retesting procedure	Yes	17d_Submission of test method based on EST-1000-epiCS for skin irritation hazard identification.pdf
13	ECVAM Test Submission Template (TST) for EST1000 SIT study / epiCS SIT study, (November 5, 2013)	Yes	ECVAM_test_submission_template_epICS_SIT_completed,05.11.2013.pdf
FOLDER: 2. DOCUMENT SET: Submission 2011			
14	Standard Operating Procedure for EPIDERMAL SKIN TEST 1000 (EST1000) SKIN IRRITATION TEST, version 1.0 of June 2011	Yes	1_INVITTOX_Protocol_EST1000_SIT_SOP.DOC
15	List of test chemicals	Yes	2_EST1000_SIT_chemicals_set.xls
16	Study Plan for the catch-up validation study on in vitro skin irritation using epidermal skin test 1000 (EST1000)- Version 3 for the initial study (2011)	Yes	13_EST1000_SIT_Study_plan.doc
17	OECD Test guideline 439 "In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method", version adopted 07/2010	Yes	15_OECD_TG_439_2010.pdf

18	Scientific publication: Hoffmann J, Heisler E, Karpinski S, Losse J, Thomas D, Siefken W, Ahr HJ, Vohr HW, Fuchs HW (2005). Epidermal-skin-test 1000 (EST-1,000)--a new reconstructed epidermis for in vitro skin corrosivity testing. Toxicol In Vitro. Oct;19(7):925-9	Yes	16_Hoffmann_EST1000.pdf
19	List of 24 references	Yes	16_References_List_EST1000_SIT.doc
20	Lot Release Certificate Epidermal Skin Test 1000 EST1000 as an example for QA	Yes	17_Lot_Release_Certificate.pdf
21	Statistical Report from BfR for the initial study: "Evaluation of the Catch-up Validation study of the Epidermal Skin Test 1000 (EST1000) for in vitro skin irritation testing", version of 14 November 2011	Yes	17_Statistical_Report.pdf
22	Summary of EST1000 SIT validation study	Yes	17_Summary_EST1000_SIT.docx
23	ECVAM Test Submission Template (TST) for the EST1000 SIT study, (15 November 2011)	Yes	ECVAM_test_submission_template_EST1000_SIT.doc
FOLDER: 3. EURL ECVAM Performance Standards			
24	ECVAM Performance Standards for Skin Irritation Testing (updated) – 24.8.2009	Yes	ECVAM_PS_2009_updated.pdf
FOLDER: 4. DOCUMENT SET: Submission 2015			
25	Excel file: trial 1 conducted in 2015 at HCCR	Yes	Harlan, trial 1, epiCS SIT method, final testing.xlsx
26	Excel file: trial 2 conducted in 2015 at HCCR	Yes	Harlan, trial 2, epiCS SIT method, final testing.xlsx
27	Excel file: trial 3 conducted in 2015 at HCCR	Yes	Harlan, trial 3, epiCS SIT method, final testing.xlsx
28	Excel file: trial 4 conducted in 2015 at HCCR	Yes	Harlan, trial 4, epiCS SIT method, final testing.xlsx
29	Excel file: trial 1 conducted in 2015 at IIVS	Yes	IIVS, trial 1, epiCS SIT method, final testing.xlsx
30	Excel file: trial 2 conducted in 2015 at IIVS	Yes	IIVS, trial 2, epiCS SIT method, final testing.xlsx
31	Excel file: trial 3 conducted in 2015 at IIVS	Yes	IIVS, trial 3, epiCS SIT method, final testing.xlsx
32	Excel file: trial 4 conducted in 2015 at IIVS	Yes	IIVS, trial 4, epiCS SIT method, final testing.xlsx
33	Excel file: trial 5 conducted in 2015 at IIVS	Yes	IIVS, trial 5, epiCS SIT method, final testing.xlsx
34	Excel file: trial 6 conducted in 2015 at IIVS	Yes	IIVS, trial 6, epiCS SIT method, final testing.xlsx

35	Excel file: trial 7 conducted in 2015 at IIVS	Yes	IIVS, trial 7, epiCS SIT method, final testing.xlsx
36	Excel file: report of the data by the submitter in the EURL ECVAM template <i>Note: EURL ECVAM offered this template to the submitter in order to facilitate consistent reporting of the data. However, EURL ECVAM specified that the use of this template was NOT mandatory and declined any responsibility on its possible inappropriate use (see disclaimer)</i>	Yes	epiCS SIT method, EURL ECVAM reporting sheet, final testing, 27.10.2015.xlsx
37	Submitter's report on the re-testing conducted in 2015	Yes	Report on epiCS SIT final testing, 29.10.2015.pdf
38	ESAC Opinion of November 2014	Yes	Ares(2014)4229618_ESAC opinion.pdf
39	EURL ECVAM Reply Letter to the Submitter of December 2014	Yes	Ares(2014)4229618_Reply letter 16122014.pdf
40	OECD Test Guideline 439 "In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method", version adopted 07/2015	Yes	OECD TG439 2015.pdf
41	OECD Series on Testing and Assessment No. 220 "Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Reconstructed Human Epidermis (RhE) Test Methods for Skin Irritation Testing as Described in TG 439", version adopted 07/2015	Yes	OECD STA220 2015.pdf

7. TERMS OF REFERENCE OF THE ESAC RAPPORTEURS

7.1 ESTABLISHMENT OF THE ESAC RAPPORTEURS

During its 38th meeting June 2013 the ESAC plenary unanimously decided to charge two ESAC members to act as rapporteurs to prepare a scientific review of a Performance Standards based study on the epiCS® RhE skin irritation test method.

7.2 TITLE OF THE ESAC RAPPORTEURS

epiCS Review 2014/2016

7.3 MANDATE OF THE ESAC RAPPORTEURS

The Rapporteurs are requested to reconsider the scientific review of the epiCS® SIT performed in 2014 on the basis of the new data and information submitted to EURL ECVAM in October 2015. The Rapporteurs are requested to revise the ESAC Rapporteur Report and Opinion of November 2014, if/where appropriate, considering the performance of the epiCS® SIT in reference to the Performance Standards on *in vitro* skin irritation testing based on RhE.

The review should focus on the appropriateness of design and conduct of the study in view of the study objective and should provide an appraisal to which extent the conclusions of the Validation Management Team (VMT) or the Submitter are substantiated by the information generated during the study and how the information generated relates to the scientific background available.

7.4 DELIVERABLES OF THE ESAC RAPPORTEURS

The ESAC WG is requested to deliver to the chair of the ESAC and the ESAC Coordinator a **revised Rapporteur Report** outlining its analyses and conclusions and a **draft revised ESAC opinion**.

The conclusions drawn in the report should be based preferably on consensus. If no consensus can be achieved, the report should clearly outline the differences in the appraisals and provide appropriate scientific justifications.

7.5 PROPOSED TIMELINES OF THE ESAC RAPPORTEURS

Item	Proposed date/time	Action	Deliverable
1	March 2014	Teleconference	Kick-off discussions
2	Beginning of April 2014	Circulation of first version of draft report	Done
3	End of April 2014	Circulation of final rapporteur report.	Done
4	End of April 2016	Circulation of first revised versions of rapporteur report and opinion	
5	End of May 2016	Circulation of final revised versions of rapporteur report and opinion	

7.6 QUESTIONS WHICH SHOULD BE ADDRESSED BY THE ESAC RAPPORTEURS

The review should address the questions put forward to ESAC by EURL ECVAM (see section 4.2) and the information requirements of the ESAC Working Group Template, where applicable. The ESAC Coordinator(s) will provide guidance if needed.

APPENDIX 1 REPORTING TEMPLATE

The appended ESAC WG template suggests a structure that is in close agreement with the EURL ECVAM information requirements ("modules") for scientific review following validation and allows at the same time for the description of the analysis and conclusions concerning more specific questions.

The template can be used for various types of validation studies (e.g. prospective full studies, retrospective studies, performance-based studies and prevalidation studies). Depending on the study type and the objective of the study, not all sections may be applicable.

However, for reasons of consistency and to clearly identify which information requirements have not been sufficiently addressed by a specific study, this template is uniformly used for the evaluation of validation studies.

The current template is

TEMPLATE_ESAC-WG_REPORT-v6.doc



EUROPEAN COMMISSION

DIRECTORATE-GENERAL

JOINT RESEARCH CENTRE

Directorate F - Health, Consumers and Reference Materials

European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Annex 3

ESAC RAPPORTEURS PEER REVIEW CONSENSUS REPORT



**EURL ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)**

ESAC Rapporteurs Peer Review Consensus Report
on the
Validation Study of the epiCS® Skin Irritation Test (SIT)
based on the EURL ECVAM/OECD Performance
Standards for *In Vitro* Skin Irritation Testing using
Reconstructed human Epidermis (RhE)

Title page information			
File name	ESAC_Rapporteurs_Report_epiCS_SIT.doc		
Abbreviated title of ESAC request	PS-based validation of epiCS® SIT		
Relating to ESAC REQUEST No.	2014-01 – updated 14/3/2016		
Request discussed through	ESAC 39 (March 2014), ESAC40 (October 2014) and by written procedure previous to ESAC 42 (June 2016)		
Report to be handed over to ESAC Chair and EURL ECVAM Coordinator by	Jon Richmond/Renate Krätke		
Version tracking			
Date	Version	Author(s)	Description
09 September 2014	V1.1	JDR	JDR draft 2014 09 09
18 September 2014	V2.0	JDR	JDR draft 2014 09 18 following receipt of response letter from CellSystems (dated 17/9) answering three questions raised by the ESAC Rapporteurs and discussion during TC with ECVAM (18/9)
26 September 2014	V2.1	RK, CG	Editorial changes following TC on 26/9
29 September 2014	V2.2	RK	Editorial changes
06 October 2014	V3.0	RK, JDR	Final Text
31 May 2016	V4.0	JDR, RK	Revised/updated version taking account of additional testing in response to 2016 request for ESAC advice
06 June 2016	V4.1	JDR, RK	Final approved draft sent to ESAC for endorsement after some final editorial changes

TABLE OF CONTENTS

TABLE OF CONTENTS	29
ESAC RAPPORTEURS	31
ABBREVIATIONS USED IN THE DOCUMENT	32
EXECUTIVE SUMMARY	33
INTRODUCTION	34
BACKGROUND	34
1. STUDY OBJECTIVE AND DESIGN	43
1.1 ANALYSIS OF THE CLARITY OF THE STUDY OBJECTIVE'S DEFINITION	43
(a) <i>Rapporteurs summary of the study objective as outlined in the Test Submission</i>	43
(b) <i>Appraisal of clarity of study objective as outlined in the Test Submission</i>	43
1.2 QUALITY OF THE BACKGROUND PROVIDED CONCERNING THE PURPOSE OF THE TEST METHOD	43
(a) <i>Analysis of the scientific rationale provided in the Test Submission</i>	43
(b) <i>Analysis of the regulatory rationale provided in the Test Submission</i>	44
1.3 APPRAISAL OF THE APPROPRIATENESS OF THE STUDY DESIGN	44
1.4 APPROPRIATENESS OF THE STATISTICAL EVALUATION	45
2. COLLECTION OF EXISTING DATA	46
2.1 EXISTING DATA USED AS REFERENCE DATA	46
2.2 EXISTING DATA USED AS TESTING DATA	46
2.3 SEARCH STRATEGY FOR RETRIEVING EXISTING DATA	46
2.4 SELECTION CRITERIA APPLIED TO EXISTING DATA	46
3. QUALITY ASPECTS RELATING TO DATA GENERATED DURING THE STUDY	47
3.1 QUALITY ASSURANCE SYSTEMS USED WHEN GENERATING THE DATA	47
3.2 QUALITY CHECK OF THE GENERATED DATA PRIOR TO ANALYSIS	47
4. QUALITY OF DATA USED FOR THE PURPOSE OF THE STUDY (EXISTING AND NEWLY GENERATED)	48
4.1 OVERALL QUALITY OF THE EVALUATED TESTING DATA (NEWLY GENERATED OR EXISTING)	48
4.2 QUALITY OF THE REFERENCE DATA FOR EVALUATING RELIABILITY AND RELEVANCE	48
4.3 SUFFICIENCY OF THE EVALUATED DATA IN VIEW OF THE STUDY OBJECTIVE	51
5. TEST DEFINITION (MODULE 1)	52
5.1 QUALITY AND COMPLETENESS OF THE OVERALL TEST DEFINITION	52
5.2 QUALITY AND COMPLETENESS OF THE DOCUMENTATION CONCERNING SOPs AND PREDICTION MODELS	52
6. TEST MATERIALS	53
6.1 SUFFICIENCY OF THE NUMBER OF EVALUATED TEST ITEMS IN VIEW OF THE STUDY OBJECTIVE	53
6.2 REPRESENTATIVENESS OF THE TEST ITEMS WITH RESPECT TO APPLICABILITY	53
7. WITHIN-LABORATORY REPRODUCIBILITY (WLR) (MODULE 2)	54
7.1 ASSESSMENT OF REPEATABILITY AND REPRODUCIBILITY IN THE SAME LABORATORY	54
7.2 CONCLUSION ON WITHIN-LABORATORY REPRODUCIBILITY AS ASSESSED BY THE STUDY	54
8. TRANSFERABILITY (MODULE 3)	55
8.1 QUALITY OF DESIGN AND ANALYSIS OF THE TRANSFER PHASE	55
8.2 CONCLUSION ON TRANSFERABILITY TO A NAÏVE LABORATORY / NAÏVE LABORATORIES AS ASSESSED BY THE STUDY	55

9. BETWEEN-LABORATORY REPRODUCIBILITY (BLR) (MODULE 4)	56
9.1 ASSESSMENT OF REPRODUCIBILITY IN DIFFERENT LABORATORIES	56
9.2 CONCLUSION ON BETWEEN-LABORATORY REPRODUCIBILITY AS ASSESSED BY THE STUDY	56
10. PREDICTIVE CAPACITY AND OVERALL RELEVANCE (MODULE 5)	57
10.1 ADEQUACY OF THE ASSESSMENT OF THE PREDICTIVE CAPACITY IN VIEW OF THE PURPOSE	57
10.2 OVERALL RELEVANCE (BIOLOGICAL RELEVANCE AND ACCURACY) OF THE TEST METHOD IN VIEW OF THE PURPOSE	58
11. APPLICABILITY DOMAIN (MODULE 6)	59
11.1 APPROPRIATENESS OF STUDY DESIGN TO CONCLUDE ON APPLICABILITY DOMAIN, LIMITATIONS AND EXCLUSIONS	59
11.2 QUALITY OF THE DESCRIPTION OF APPLICABILITY DOMAIN, LIMITATIONS, EXCLUSIONS	59
12. PERFORMANCE STANDARDS (MODULE 7)	60
12.1 ADEQUACY OF THE PROPOSED ESSENTIAL TEST METHOD COMPONENTS	60
12.2 ADEQUACY OF THE REFERENCE CHEMICALS	60
12.3 ADEQUACY OF PROPOSED PERFORMANCE TARGET VALUES	60
13. READINESS FOR STANDARDISED USE	61
13.1 ASSESSMENT OF THE READINESS FOR REGULATORY PURPOSES	61
13.2 ASSESSMENT OF THE READINESS FOR OTHER USES	61
13.3 CRITICAL ASPECTS IMPACTING ON STANDARDISED USE	61
13.4 GAP ANALYSIS	61
14. OTHER CONSIDERATIONS	62
15. CONCLUSIONS ON THE STUDY	63
15.1 ESAC WG SUMMARY OF THE RESULTS AND CONCLUSIONS OF THE STUDY	63
15.2 EXTENT TO WHICH STUDY CONCLUSIONS ARE JUSTIFIED BY THE STUDY RESULTS ALONE	63
15.3 EXTENT TO WHICH CONCLUSIONS ARE PLAUSIBLE IN THE CONTEXT OF EXISTING INFORMATION	63
16. RECOMMENDATIONS	64
16.1 GENERAL RECOMMENDATIONS	64
16.2 SPECIFIC RECOMMENDATIONS (E.G. CONCERNING IMPROVEMENT OF SOPs)	64
17. REFERENCES	65

ESAC Rapporteurs

Rapporteurs were appointed by the ESAC 39 meeting to facilitate the ESAC Opinion requested in the “ECVAM REQUEST FOR ESAC ADVICE on the validation study of the epiCS® test method based on the EURL ECVAM/OECD Performance Standards for *in vitro* skin irritation testing using *Reconstructed human Epidermis* (RhE)” dated 21 February 2014 (ESAC Request 2014-01).

The Rapporteurs were tasked with conducting a detailed scientific peer review of an external Performance Standards-based “catch-up” validation study submitted by the test method developer based on the provisions of the relevant Performance Standards (PS) (EURL ECVAM Performance Standard 2009, and OECD TG439 2013) for *in vitro* skin irritation test methods using *Reconstructed human Epidermis* (RhE).

The October 2014 Rapporteurs Report (V3) represented their consensus view on the information available at that time.

The resulting ESAC Opinion (2014-1) of 17/11/2014 recommended that supplementary testing should be conducted at two laboratories to remove potential sources of bias from the chemical testing datasets.

The Rapporteurs have now been asked (ESAC Request 2014-01, updated 31/03/2016¹) to take account of additional data and analysis provided in a 2015 Report from the test developer in response to the ESAC Opinion and subsequent correspondence.

This document (V 4.0) sets out the Rapporteurs’ agreed position taking this additional information in to account.

The ESAC Rapporteurs:

- Dr. Jon Richmond
- Dr. Renate Krätke

ESAC Coordination:

- Dr. Claudius Griesinger (ESAC Coordinator; 2014)
- Dr. Michael Schäffer (2014)
- Dr. João Barroso (ESAC Coordinator; 2016)

¹ The revised ESAC request references the Performance Standards set out in the 2015 OECD documents (the original ESAC requested made reference to the 2013 Edition).

OECD (2015a). OECD Series on Testing and Assessment, No. 220: Performance Standards for the Assessment of Proposed Similar or Modified *In Vitro* Reconstructed Human Epidermis (RhE) Test Methods for Skin Irritation Testing as Described in TG 439. Organisation for Economic Cooperation and Development, Paris.

OECD (2015b). OECD Guideline for the Testing of Chemicals, No. 439: *In Vitro* Skin Irritation: Reconstructed Human Epidermis Test Method. Organisation for Economic Cooperation and Development, Paris.

Abbreviations used in the document

• BLR	Between-laboratory reproducibility
• EURL ECVAM	European Union Reference Laboratory for Alternatives to Animal Testing
• ESAC	EURL ECVAM Scientific Advisory Committee
• ESAC WG	ESAC Working Group
• GLP	Good Laboratory Practice
• OECD	Organisation for Economic Cooperation and Development
• PC	Positive Control
• PS	Performance Standards
• RC	Reference Chemical
• RhE	Reconstructed human Epidermis
• SD	Standard deviation
• SOP	Standard Operating Procedure (used here as equivalent to 'protocol')
• TS	Test Submission
• VMT	Validation Management Team
• VRM	Validated Reference Methods
• WLR	Within-laboratory reproducibility

EXECUTIVE SUMMARY

In 2014 ESAC was requested to peer review and offer an opinion as to whether the performance of the epiCS® RhE *in vitro* skin irritation method satisfies the relevant Performance Standards with a view to consideration of its use for routine dermal irritation testing hazard classification for regulatory purposes.

The validation study was set up and managed by the test developer. The initial ring-trial was conducted in 2011. Further testing was performed in 2013 as the result of suspected proficiency problems at the testing laboratories (see 1.3.2 and 8). Additional testing was conducted and reported in 2015 following a 2014 ESAC Recommendation raising concerns about possible sources of bias in the data-set made available to ESAC in 2014 (see 1.3.3 and 1.4 and 4.2).

Having considered the 2015 test developer's Report to EURL ECVAM, the ESAC Rapporteurs requested additional analysis in order to derive a dataset that they consider a more appropriate way to judge the true performance of the test method (see below and 10). The results which the ESAC Rapporteurs believe should be used to judge the performance of the test method are shown in table 1 immediately below, including Clopper-Pearson exact 95 %-Confidence Intervals calculated by ESAC.

TABLE 1: ESAC Rapporteurs' Re-calculated FINAL Results

Validation Study Results based on Rapporteurs' calculations			PS Criteria Met Yes/No
Lab. #1 – ACS	Sensitivity	100 % (66–100 %)	
	Specificity	80 % (44–97 %)	
	Accuracy	89 % (67–99 %)	
Lab. #2 – HCCR	Sensitivity	90 % (56–100 %)	
	Specificity	80 % (44–97 %)	
	Accuracy	85 % (62–97 %)	
Lab. #3 – IIVS	Sensitivity	75 % (35–97 %)	
	Specificity	80 % (44–97 %)	
	Accuracy	78 % (52–94 %)	
For all three laboratories	Sensitivity	89 % (71–98 %)	Yes (PS ≥ 80 %)
	Specificity	80 % (61–92 %)	Yes (PS ≥ 70 %)
	Accuracy	84 % (72–93 %)	Yes (PS ≥ 75 %)
WLR Lab. #1 – ACS		95 % (74–100 %)	Yes (PS ≥ 90 %)
WLR Lab. #2 – HCCR		100 % (83–100 %)	Yes (PS ≥ 90 %)
WLR Lab. #3 – IIVS		89 % (65–99 %)	No (PS ≥ 90 %)
BLR		88 % (64–99 %)	Yes (PS ≥ 80 %)
% of complete run seq. in Lab. #1 – ACS		95 % (75–100 %)	Yes (PS ≥ 85 %)
% of complete run seq. in Lab. #2 – HCCR		100 % (83–100 %)	Yes (PS ≥ 85 %)
% of complete run seq. in Lab. #3 – IIVS		90 % (68–99 %)	Yes (PS ≥ 85 %)
% of complete run seq. over the three labs		95 % (86–99 %)	Yes (PS ≥ 90 %)

With the exception of the WLR in one laboratory (at 88.89%, marginally short of the PS required minimum value of ≥ 90%) all of the PS criteria are met.

Whilst one of the relevant Performance Standards makes provision for deviations from the acceptance criteria² - the other does not.

INTRODUCTION

In 2014 ESAC was requested to evaluate a study plan and a composite data set (submitted by the test developer to ECVAM in 2013) selectively merging and substituting data from two phases of chemical testing: an initial 2011 study (TM2009-09), and a later testing phase generating supplementary test data on selected chemicals after additional training and proficiency testing measures were introduced at the participating laboratories (see 1.3.2 and 8). ESAC was specifically asked to review and offer an opinion as to whether the overall performance of the epiCS® *in vitro* skin irritation method satisfies the relevant Performance Standards in view of its potential use for routine dermal irritation testing hazard classification for regulatory purposes.

The resulting ESAC Opinion (2014-01 (17/11/2014)) concluded on the basis of that hybrid data-set that the majority of the PS performance criteria had been met. However, there were concerns that the testing programme and hybrid data-set produced by selectively mixing test data from 2011 and 2013 may have introduced an “optimism bias” making it impossible at that time to properly evaluate test method performance (see 1.3.3 and 1.4 and 4.2): the ESAC opinion therefore recommended further testing and analysis to address this concern.

That resulted in a third round of testing which forms the basis of a Report from the test developer to EURL ECVAM in October 2015. ESAC has now been asked to advise on the performance of the test method taking account of data and analysis presented in the test developer’s Report to ECVAM of October 2015.

BACKGROUND

2009

CellSystems submitted an ECVAM Test Method "Proposal Evaluation Form", providing a short outline and description of the EST-1000 SIT *in vitro* skin irritation test method. ECVAM informed the submitter that the test method may qualify for a Performance Standards-based validation study subject to a more complete evaluation confirming sufficient methodological similarity between the EST-1000 assay and the validated reference methods (VRMs). The submitter was invited to submit additional information.

2010

A more detailed test method description was submitted by the test developer in 2010 addressing the essential test method components outlined in the ECVAM PS for *in vitro* skin irritation test methods based on Reconstructed human Epidermis (RhE) test systems (ECVAM 2009). The assessment then performed by ECVAM indicated **full compliance with these criteria** and ESAC (in Q1 2011) agreed with this assessment. ECVAM informed the submitter that the test method would qualify for PS-based validation and invited the submitter to submit a revised Test Submission (TS) once a PS-based ring trial employing three laboratories had been completed.

² For example, the 2009 ECVAM Performance Standard (page 5) states “...it is conceivable that deviations from these standards may be justified by scientific reasons...”. The OECD Performance Standard does not contain this provision.

2011

In 2011 the test developer, having conducted a first PS-based validation study of the epiCS® test method (then known as EST-1000), submitted the findings to EURL ECVAM. While the performance indices derived from the study data met the target values relating to Predictive Capacity (Specificity, Sensitivity and Accuracy) as outlined in the PS, the trial data did not satisfy the criteria established in the PS for Reproducibility: within Laboratory Reproducibility (WLR) being 16% below the target value in the naïve laboratory, and 6% below the target value in the other two laboratories; and the Between Laboratory Reproducibility (BLR) being 2% below the target value.

The 2011 ring-trial did not include a transfer, training and proficiency phase (see 1.3.2 and 8). It was considered that this omission may have contributed to the poor Reproducibility values reported in the 2011 submission. No ESAC opinion was sought at that time.

The Rapporteurs note that during this phase of the validation study the nature of the plastic inserts which form part of the test kit were changed, but that this did not compromise test performance values. The Rapporteurs consider that the ability of the test method to withstand such a change in manufacturing process tends to confirm the robustness of the test method and its external validity.

2013

After remedial action by the manufacturer to improve the transport of the test kits to the laboratories, and to improve the technical proficiency and compliance of the laboratories in particular with respect to how to execute crucial SOP steps (relating to chemical storage, test material application and its removal, i.e. rinsing procedure), supplementary testing of the full set of RCs at the naïve laboratory, and re-testing of a subset of chemicals at the other laboratories was undertaken.

The supplementary testing and analysis that under-pinned the 2013 submission can be summarised as follows: the naïve laboratory had the worst WLR performance and re-tested the original 20 reference chemicals; the other two laboratories had also failed to meet the WLR acceptance criteria as outlined in the PS (which states 90% of the 20 chemicals to be concordantly predicted between runs) and re-tested only the six chemicals responsible for most of the Reproducibility errors observed, and a seventh chemical in one of the laboratories only because it generated invalid run sequences during the 2011 ring-trial. In the 2013 submission to EURL ECVAM revised WLR and BLR values had been calculated by substituting the new data from all three laboratories for the results obtained for the same chemicals in the same laboratories as reported in the 2011 submission to EURL ECVAM. It is on that basis revised Reproducibility and Accuracy figures were calculated and supplied to EURL ECVAM by the test developer in 2013.

In their 2014 peer review the Rapporteurs were satisfied that the available data tended to confirm that the most plausible explanations for the poor Reproducibility in the 2011 ring trial were:

- (a) The poor proficiency, in particular but not exclusively, of the naïve laboratory. Moreover, on the basis of 2013 supplementary data, and additional analysis provided by the test method developer in a letter dated September 2014, this was also seen to have been an issue at the two non-naïve laboratories. Proficiency and compliance issues mainly concerned the washing/rinsing procedure, which, although described in the SOP, was not properly followed.
- (b) Problems with the shipment of the tissues in case of the US laboratory.
- (c) The fact that tissue kits that had been exposed to different chemicals had been placed next to each other during the post-incubation period. While this step follows the washing procedure intended to remove the test chemical, potential cross-contamination issues could not be ruled-out (see also section 16. Recommendations).

In evaluating the 2013 submission to EURL ECVAM, the Rapporteurs considered that elements of the supplementary testing and data analysis may have introduced sources of bias into the datasets submitted to EURL ECVAM in 2013 (as discussed in more detail below and summarised at 1.3.3 and 1.4), and to have breached the PS requirements with respect to repeated testing. The Rapporteurs therefore viewed the 2013 submission as an adjunct explaining and supporting, rather than replacing, the 2011 submission.

The Rapporteurs were therefore concerned that the reported Reproducibility and Accuracy of the epiCS® test method set out in the 2013 Test Submission may not be based on the most reliable dataset to use to interpret the findings of this study with respect to Reproducibility or Reliability. Specifically, the Rapporteurs concluded that there were two features of the 2013 study design and data analysis which represented potential sources of bias potentially over-estimating the performance of the test method.

- First: the PS state that for validation studies, the Reference Chemicals (RCs) which are used for PS-based validation, should not be used for development and optimisation of the test method (EURL ECVAM, 2009; OECD, 2015a). However, to develop and optimise the epiCS® test method 44 chemicals had been used, including the majority of the 20 RCs listed in the PS (see 1.3.1, and 2.4 and 4.2 and 6).
 - However, the Rapporteurs also noted at that time that the PS provision not to use RCs for test method development and optimisation had not been observed with respect to other “me-too's” previously validated, i.e. all other validated RhE methods also used the majority of RCs for test method development and optimisation.
- Second: with respect to the supplementary testing conducted by the two “non-naïve” laboratories, the Rapporteurs felt that this was not the appropriate approach (see 1.3.3 and 1.4 and 4.2 and 10): interpreting the 2011 and 2013 data sets would have been considerably easier had the full set of chemicals been re-tested by all laboratories and not only by the naïve laboratory. The relevant PS (see for example OECD STA 220, 2015a) regard post-hoc data selection as inappropriate, and the 2013 submission included post-hoc data selection.

The Rapporteurs considered at that time that a more appropriate, but still imperfect, dataset on which to present and judge to the likely true performance of the epiCS® method using the data available from this validation study was to combine and analyse the 2013 data from only the naïve laboratory which re-tested all 20 reference chemicals (believing that this element of the supplementary testing did address the most plausible avoidable errors introduced by the initial failure to incorporate formal training and proficiency testing and was justified) with the original 2011 datasets from the other two laboratories (which had re-tested only a carefully selected subset of the test chemicals – that is, those responsible for most of the reproducibility errors in the 2011 study).

Following this approach the performance parameters which the Rapporteurs’ at that time considered the most appropriate to judge the performance of the test were as follows (see 10):

- The re-calculated WLR values produced on this basis were 95%, 84% and 84%. The minimum acceptable value in the PS is 90%.
- However, when considering the figures presented in the 2013 Test Submission including also the new data from the non-naïve laboratories, were 95%, 100% and 95%, tending to confirm that with experience and full compliance with the SOP acceptable WLR values can be obtained.

TABLE 2: WLR VALUES (PS target: WLR ≥ 90% concordant predictions)

LABORATORY	2011 Data	2013 Data	ESAC WG Re-calculation	COMMENT
ACS (NAIVE)	74%	95%	95%	Re-calculation tended to confirm that poor proficiency contributed to the poor WLR values in the 2011 report. Another confounding aspect may have been cross contamination during the post-treatment incubation step.
HCCR	84%	100%	84%	Re-calculation tended to confirm that even experienced laboratories may require training and proficiency testing to transfer the protocol. Another confounding aspect may have been cross contamination during the post-treatment incubation step.
IIVS	84%	95%	84%	Re-calculation tended to confirm that even experienced laboratories may require training and proficiency testing to transfer the protocol. Another confounding aspect may have been cross contamination during the post-treatment incubation step. Moreover, it is plausible that shipment issues may have negatively impacted on the tissue quality of the 2011 study.

- The re-calculated Between Laboratory Reproducibility (BLR) value was 88% (with consistent predictions for 15 chemicals out of 17 chemicals with valid predictions in all three laboratories). The minimum acceptable value in the PS is 80%. The equivalent figure in the 2013 Test Submission (considering all adjunct testing) was 95%.

TABLE 3: BLR VALUES (PS target: BLR ≥ 80% concordant predictions)

2011 Data	2013 Data	ESAC WG Re-calculation	COMMENT
78%	95%	88%	Tended to confirm the need for training and proficiency testing in laboratories that wish to establish the test method.

- With respect to Accuracy, the values re-calculated by the Rapporteurs were: Sensitivity 95%, Specificity 79%, and overall Accuracy 87%. These values satisfied the acceptance values set out in the Performance Standards (80%, 70%, and 75%).

2014

In 2014 the Rapporteurs reasoned that the ESAC Opinion and EURL ECVAM decision must be based on what can be inferred about the values that may have been achieved had the original study design not been compromised by a failure to implement and transfer/training/proficiency phase (see 1.3.2), and had the full set of test chemicals been re-tested by all three laboratories after the potential proficiency issues been resolved. However, the basis of the calculations in the 2013 submission may have introduced elements of bias over-estimating the performance of the test method (see above and 1.3.3 and 1.4 and 10).

The Rapporteurs' findings informed the resulting ESAC Opinion 2014-01 of 17 November 2014:

“Recommendations

At its 40th plenary meeting in October 2014, ESAC discussed the study design and data sets submitted in 2011 and in 2013 and concluded that Performance Standard acceptable values for BLR, Specificity, Sensitivity and Overall Accuracy were met in the validation study. However, only the naïve laboratory satisfied the performance values relating WLR when re-testing all 20 reference chemicals in 2013. The two other laboratories only met performance values when mixing data from the 2011 testing with data from the re-testing in 2013, generated under different testing conditions. ESAC is of the opinion that mixing of data from 2011 and 2013 may introduce sources of bias into the data set submitted to EURL ECVAM in 2013. Therefore values for WLR for these two laboratories can only be derived from the 2011 data set and are considered to be below the acceptance threshold, although evaluation of all data together tends to support that performance values for WLR can be attained. For the final evaluation of the WLR, ESAC recommends to re-test also the remaining 13 and 14 chemicals so that, ultimately, the full set of chemicals is available also from laboratory 2 and laboratory 3 in view of assessing the final WLR values and judging whether these meet the Performance Standard acceptable value of 90%.

ESAC also suggested amendments to the SOP to stress:

- (1) the importance of an accurate rinsing procedure;*
- (2) to improve the description of the washing procedure in the SOP;*
- (3) that tissues treated with different test chemicals should not be placed next to each other neither during exposure nor during post-incubation.”*

2015

In October 2015, in response to ESAC Opinion 2014-01, the manufacturer submitted a Report³ to EURL ECVAM claiming compliance with ESAC's 2014 recommendations with respect to the need for additional test data providing new data for the 13 and 14 chemicals that had not been re-tested respectively by HCCR and IIVS in the second validation trial conducted in 2013 (Table 4).

That Report stated that the two laboratories had also undergone another round of training and proficiency testing before generating the additional data. For this they used 5 chemicals from the OECD PS table of Proficiency Substances: however, all 5 of these chemicals were also on the list of RCs the laboratories then re-tested. In the view of the Rapporteurs a different set of chemicals should have been used for training and proficiency testing in this case (see 1.1.4, see 2.4 and 6).

³ “Report on epiCS® Skin Irritation Test method Validation Study - final testing and submission”, date: October 29, 2015. Author: Dr Oliver Engelking, Cell Systems Biotechnologie Vertrieb GmbH, Troisdorf, Germany.

TABLE 4: Different phases of the PS-based validation of the epiCS® SIT resulting in all three laboratories re-testing all chemicals after training and proficiency testing – as reported by the test method developer

ACS	HCCR	IIVS
2011		
20 chemicals	20 chemicals	20 chemicals
2013 + 2015		
2013	2013	2013
20 chemicals re-tested	7 chemicals re-tested	6 chemicals re-tested
	2015	2015
	13 chemicals re-tested	14 chemicals re-tested

The results reported in the 2015 submission with the pooled dataset from the studies of 2013 and 2015 as recommended by ESAC (with all 20 chemicals tested in all three laboratories after training and proficiency testing) can be summarised as:

- **Occurrence of non-qualified tests/runs:**
 - If values for the positive or negative controls were outside of the accepted ranges ($\text{Viability}_{\text{PC}} \leq 20\%$; $1.0 \leq \text{OD}_{\text{NC}} \leq 2.8$) the run was considered not qualified and, consequently, all tests included within that run were also considered not qualified.
 - As per the PS the mean of triplicate values of each control and test was calculated and if the SD was $> 18\%$, the run and/or test was considered not qualified.
 - Thirty four out of the 212 (16 %) tests conducted in 2013 + 2015 were non-qualified. Of these 34 non-qualified tests, 16 occurred due to failure of the positive control rather than the SD of the test being $> 18\%$.
 - The incidence of non-qualified tests is similar to what was observed in the original study of 2011 (18 % non-qualified tests).
 - The study acceptance criteria as stipulated in the PS with respect to these elements of the validation study were reported as having been met.
- **Complete run sequences:**
 - The aggregated data from two of the three laboratories were reported as having complete run sequences for all 20 chemicals following re-testing (100 %) (but see below and 1.3.5 and 1.4 and 10).
 - The third laboratory, the naïve laboratory (ACS), was reported as having complete run sequences for 19 out of 20 chemicals (95 %), thus meeting the PS acceptance criterion of $\geq 85\%$ (based on data from 2013).
 - The incomplete run sequence was obtained for 1-methyl-3-phenyl-1-piperazine; ACS was unable to generate three qualified runs after the PS permitted maximum of 5 independent runs.

- The 2013 manufacturer's report did not reference incomplete run sequences at the other two laboratories.
- In the 2015 test manufacturer's Report to EURL ECVAM 98.3 % of the run sequences were reported as complete, seeming to satisfy the acceptance criterion of ≥ 90 % (but see below and 1.3.5 and 1.4 and 4.2 and 10).
- The reported **WLR of 94.7 % at ACS, 100 % at HCCR and 90 % at IIVS** meets the acceptance criterion of ≥ 90 % (but see below and 1.3.5 and 1.4 and 4.2 and 10).
- The reported **BLR of 84.2 %** (based on results from 19 RCs) meets the acceptance criterion of ≥ 80 % (but see below and 1.3.5 and 1.4 and 4.2 and 10).
 - The PS (OECD 2015a) requires that only chemicals with three complete test run sequences are used for the BLR calculation.
 - Since the reference chemical 1-methyl-3-phenyl-1-piperazine did not have complete run sequences in all laboratories, it was excluded by the manufacturer for the calculation of BLR.
 - The Rapporteurs note that a qualified false negative result and a qualified true positive result were obtained for this chemical at HCCR and IIVS, respectively. Therefore, the chemical did not show reproducible results between laboratories notwithstanding of the incomplete run sequence obtained by ACS. If this chemical had been considered in the calculation of BLR, the value obtained would be 80 %.
 - The 2015 manufacturer's report did not reference incomplete run sequences for any chemical at the other two laboratories (but see below and 1.3.5 and 1.4 and 4.2 and 10).
- The reported **Sensitivity of 86.2 %** (100 % at ACS, 90 % at HCCR, 70 % at IIVS) meets the acceptance criterion of ≥ 80 % (but see below and 1.3.5 and 1.4 and 4.2 and 10).
- **False negative results** (based on the mean viability of the three qualified runs) as reported were:
 - For 1-methyl-3-phenyl-1-piperazine at HCCR.
 - For 1-methyl-3-phenyl-1-piperazine ACS produced 1 qualified true positive result and 3 non-qualified false negative results, with the mean of the four runs also being false negative.
 - For 1-bromohexane, 1-decanol and di-n-propyl disulphide at IIVS.

The manufacturer stated the acceptance criteria for false negative results specified in the PS are met.

- The reported **specificity of 80 %** (80 % in all participating laboratories) meets the PS acceptance criterion of ≥ 70 % (but see below and 1.3.5 and 1.4 and 4.2 and 10).
- **False positive results** (based on the mean viability of the three qualified runs) were reported for allyl-phenoxy-acetate and cinnamaldehyde in all three participating laboratories. The manufacturer claimed this satisfies the provisions of the PS.
- On the basis of the reported **overall accuracy of 83.1 %** (89.5 % at ACS, 85 % at HCCR, 75 % at IIVS) the manufacturer claims this meets the acceptance criterion of ≥ 75 % (but see below and 1.3.5 and 1.4 and 4.2 and 10).

2016

In reviewing the 2015 Report submitted by the test developer to EURL ECVAM and all the other available documentation and test data, the Rapporteurs noted a significant deviation from the PS with respect to the maximum permissible number of re-tests not set out in the 2015 test developer Report to EURL ECVAM (see 1.3.5): some non-qualified runs had been not tabulated in the EURL ECVAM template, and had not been taken into account when the performance indices provided in the 2015 manufacturer's Report to EURL ECVAM were calculated (see 1.3.5 and 1.4 and 4.1 and 4.2 and 10).

- The PS specify that *"to complement missing data, a maximum of two additional runs... ("re-testing")... may be conducted for each Reference Chemical in each laboratory. Non-qualified tests should be documented and reported"* (OECD 2015a).
- This means that each laboratory must not test the same reference chemical more than five times, and all non-qualified tests must be taken into account.
- The PS make no provision for disregarding or discounting test runs known to be non-qualifying due to technical reasons.
- However reviewing the raw data for the supplementary testing conducted at IIVS the Rapporteurs noted 7 test runs were required to produce three qualifying test runs for 1-decanol; and 6 test runs were required to produce three qualifying test runs for potassium hydroxide (5% aq).
- Some of these extra runs (documented in the full data-set) were not reported in the EURL ECVAM reporting template as "not qualified".
- Within IIVS "Trial 2" tests were non-qualifying due to failure of the PC, and in "Trial 5" there were reported problems with bacterial/fungal contamination and failure of the positive control.
- A true positive result (with SD < 18 %) was nevertheless obtained for 1-decanol in "Trial 2". If this result had been accepted, this chemical would have produced non-concordant results within IIVS, thus lowering the laboratory's WLR to 85 % (PS acceptance criterion of ≥ 90 %).

The Rapporteurs are therefore not prepared to rely on the analysis and indices provided in the 2015 manufacturer's Report to EURL ECVAM to offer an opinion about the true performance of the test method when measured against the PS (see 10).

The Rapporteurs believe that in order to judge the test method performance against the PS then:

- Both chemicals (1-decanol and potassium hydroxide (5% aq)) which were re-tested at IIVS using more test runs than the PS permits to produce complete run sequences must be considered to have failed to produce three qualified run sequences at this laboratory.
- The BLR must therefore be re-calculated on the basis of the 17 chemicals for which there are complete run sequences in all three laboratories.
- The WLR, Predictive Capacity and other PS required performance figures must also be re-calculated on this basis.

When that it is done the results which the Rapporteurs believe should be used to judge the performance of the test method are shown in Table 1 (see Executive Summary).

With the exception of the WLR in one laboratory (88.89%, marginally short of the PS required minimum value of ≥ 90 %) all of the PS criteria are met.

Whilst one of the relevant Performance Standards makes provision for deviations from the acceptance criteria⁴ - the other does not.

⁴ For example, the 2009 ECVAM Performance Standard (page 5) states “...it is conceivable that deviations from these standards may be justified by scientific reasons...”. The OECD Performance Standard does not contain this provision.

1. Study objective and design

In requesting advice from ESAC in 2014, EURL ECVAM advised that the Test Submissions for the validation study (an initial formal Test Submission in 2011 which was not put forward to ESAC peer review, and a revised Test Submission in 2013 following trouble-shooting, supplementary testing and data analysis) had been designed and conducted with the specific objective of assessing whether the data generated in support of the reproducibility and relevance of the epiCS® test to determine the skin irritation potential of chemicals satisfied the requirements of the relevant PS in terms of Reproducibility within and between laboratories and with respect to Accuracy of predictions (Sensitivity, Specificity, and Overall Accuracy).

In the subsequent 2016 revised request for advice ESAC was asked to take account of the 2015 Report from the test method developer and the then current PS.

1.1 Analysis of the clarity of the study objective's definition

(a) Rapporteurs summary of the study objective as outlined in the Test Submission

The documentation is clear that the study objective (c.f. Test Plan epiCS SIT, version 2011 and version 2013, and 2015 Report from the test developer) was to show that the *in vitro* RhE epiCS® test method (initially trademarked and evaluated as EST-1000) satisfactorily discriminates between non-classified ("no category") and classified ("category 1" or "category 2") chemicals for skin corrosion/irritation according to the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS. UN, 2015). The test method is not intended to differentiate between category 1 (skin corrosion) and category 2 (skin irritation).

To realise this objective, a PS-based ("catch-up") validation study based on the relevant PS (ECVAM 2009 Performance Standards for *in vitro* skin irritation test methods based on *Reconstructed human Epidermis* (RhE), and the Performance Standards in Annex 2 of OECD TG 439 (OECD, 2013) now published separately in the OECD Series on Testing and Assessment (OECD, 2015a)) was planned and conducted involving three laboratories using coded chemicals for blind testing followed by independent statistical analysis by statisticians of the federal Institute for Risk Assessment (BfR, Germany) in case of the 2011 study and, in case of the adjunct 2013 study, a statistical consultant (SHE consulting, Paderborn, Germany). Supplementary test data and analysis are set out in the 2015 Report from the test developer.

(b) Appraisal of clarity of study objective as outlined in the Test Submission

The study objectives are sufficiently clear: however, the wording indicates that the objectives relate not to determining *whether* the test method has these properties, but to seeking confirmation that it does.

1.2 Quality of the background provided concerning the purpose of the test method

A PS-based "catch-up" validation study was conducted with the implied purpose of determining whether the epiCS® test method is considered suitable for regulatory use in the same context as the RhE VRMs from which the PS were derived.

(a) Analysis of the scientific rationale provided in the Test Submission.

Scientific rationale is provided in the 2011 and 2013 Test Submissions to EURL ECVAM. These clearly describe the biological and mechanistic relevance on the epiCS® test method (known as EST-1000 at the time of the 2011 Test Submission) in terms which seek to establish that it meets the essential test

method requirements set out in the EURL ECVAM Performance Standard (2009) and the OECD TG 439 (2013⁵) relating to RhE models for assessing skin irritation potential.

In 2009 EURL ECVAM (then ECVAM) accepted that the test method satisfied the relevant essential test method requirements of the then current PS and that it hence qualified for a PS-based validation study using, as a minimum, a pre-defined set of 20 RCs to assess test performance and WLR, BLR and predictive capacity.

(b) Analysis of the regulatory rationale provided in the Test Submission

This information can be derived from the study documents and the PS referenced in the study documents, and are consistent with the provisions of the PS and testing requirements.

1.3 Appraisal of the appropriateness of the study design

The PS provide the general framework for study design and data analysis. However there are a number of problems with the way this programme of work was planned, conducted, analysed and reported that complicated the Rapporteurs' attempts to determine the true test method performance in line with the PS provisions and requirements.

In the view of the Rapporteurs the most serious shortcomings with the study design in relation to the provisions of the Performance Standards are:

1. **The selection of RCs for the validation study does not reflect best practice with respect to the provisions of the PS.** The 20 RCs used for the study are the minimum base-set as specified in the PS (EURL ECVAM, 2009; OECD, 2015a). As such they are considered suitably representative and balanced with respect to different chemical classes and physical states, and skin irritation potential as determined by the best available reference data. However, the PS recommend that chemicals used to develop and optimise test methods should not then be used for validating the test methods; and make provision for the use of other chemicals in their place. In this case the majority of the OECD listed RCs had been used to develop and optimise the test method. See also 2.4.
 - a. However, the Rapporteurs are aware that this deviation from the provisions of the PS is also a feature of all of the validation studies for other "me too" methods of this class that are now included in TG 439 (see for example Tornier et al., 2010).
2. **No training/transfer and proficiency phase was included in the initial ring-trial** - with poor proficiency then seeming to be a plausible explanation for the variability seen in the 2011 dataset. The Rapporteurs are concerned that this failure to incorporate formal transfer, training and proficiency testing phases within the ring-trial which formed the basis of the 2011 Test Submission set the scene for further problems as steps were then taken to generate a data set permitting a better evaluation of the true performance of the test method by proficient laboratories
3. **As a result of selective re-testing and selective data-substitution the test data and analysis in the 2013 submission to EURL ECVAM may have introduced sources of "optimism bias" with respect to the potential performance of the test method.** In 2014 the Rapporteurs felt that the selective supplementary testing and statistical analysis which are incorporated in the 2013 Test Submission, may have introduced sources of bias which, when taken together with the concerns about the choice of RCs, made it impossible at that time to properly to determine the most relevant, representative and reliable values to use to judge the reproducibility and reliability of the test method.

⁵ Now superseded by 2015 editions (OECD, 2015a), which do not change the essential test method requirements or the minimum required performance values.

4. **Between 2013 and 2014 the two laboratories which were to do the supplementary testing underwent additional training and proficiency testing using 5 of the chemicals which they then went on to re-test to provide data for the 2015 test method manufacturer's Report to EURL ECVAM:** the supplementary training and proficiency testing should not have used any of the chemicals which were to be re-tested.
5. **Data generated to produce the 2015 manufacturer's Report to EURL ECVAM (but not included or used for data analysis within that Report) breached the PS' provisions with respect to permissible re-testing to produce qualified test runs: some of the new data was excluded from the completed EURL ECVAM reporting template, and from the test manufacturer's data analysis and conclusions about test method performance.** This has necessitated the Rapporteurs commissioning additional data analysis in an attempt to produce a data-set that that does comply with the re-testing provisions of the PS.

The inclusion of a naïve test laboratories is in line with the provisions of the PS. Using the PS as the benchmark, the Rapporteurs can confirm that the number of participating laboratories, the number of RCs, the use of PCs and NCs, and the criteria for test run acceptance meet the requirements for a PS-based "catch-up" validation study of this nature.

1.4 Appropriateness of the statistical evaluation

The 2011 validation study meets the provisions of the Performance Standards with respect to statistical power, number of test runs and replicates, and, subject to the comments immediately below, the computational rules followed to calculate reproducibility and reliability as reported.

As Rapporteurs we have concerns regarding:

- The nature of the selective supplementary testing undertaken after consideration of the 2011 Test Submission data, and the way that data was then selectively substituted for the original 2011 test data in the 2013 submission to EURL ECVAM: both may have introduced sources of bias which must be borne in mind when evaluating the reproducibility and reliability values, and other test performance indices, reported in the 2013 Test Submission.
- The exclusion of some non-qualified test runs from the 2015 manufacturer's report to EURL ECVAM, and the manufacturer's failure to take account of these in the data analysis.

In forming a reasoned opinion on the true test method performance, the Rapporteurs commissioned additional data analysis to correct for this last shortcoming.

2. Collection of existing data

2.1 Existing data used as reference data

This epiCS® validation study was conducted using the PS listed 20 RCs. The study documents provide no additional reference data.

2.2 Existing data used as testing data

This epiCS® validation study was conducted using the 20 RCs listed in the relevant PS, and attempted to satisfy the provisions of the PS with respect to the minimum acceptable values for reproducibility and reliability.

The study documents provide no additional reference data.

2.3 Search strategy for retrieving existing data

No systematic search strategy is set out in the documents supplied by the test kit manufacturer.

2.4 Selection criteria applied to existing data

The test developer had the option of adjusting the test chemical set to take account of the chemicals previously used to develop and optimise the test method by substituting or including additional chemicals, but did not do so. This decision is not discussed, explained, or justified in the study documents.

For the supplementary training and proficiency test at the two laboratories which provided the additional data for the 2015 manufacturer's Report to EURL ECVAM the test developer had the option of using a chemical set which did not contain any of the chemicals which were to be re-tested, but did not do so. This decision is not discussed, explained, or justified in the study documents.

3. Quality aspects relating to data generated during the study

3.1 Quality assurance systems used when generating the data

The manufacture of the epiCS® test kit is undertaken according to defined quality assurance procedures, and Lot Release Quality Control and Certification confirm essential morphological and functional properties before test kits are issued for use by the manufacturer.

The 2015 manufacturer's Report to EURL ECVAM records that one non-qualifying test run at IIVS was due to bacterial/fungal infection: the study documents do not provide any insights as to whether this was a problem with the test kit as supplied by the manufacturer, or contamination at the test laboratory. The study documents also do not specify in which tissues the contamination occurred, i.e. the NC treated tissues, the PC treated tissues and/or the test chemical treated tissues.

Three laboratories conducting the testing summarised in the manufacturer's submissions undertook the studies "in the spirit of GLP", but had not been required to demonstrate proficiency with the test method before the initial ring-trial got underway.

The 2015 manufacturer's Report to EURL ECVAM mis-coded some non-qualified test runs due to possible technical issues as "not valid" rather than non-qualified.

3.2 Quality check of the generated data prior to analysis

This validation study was undertaken and co-ordinated by the test manufacturer: it was not initiated, owned or managed by EURL ECVAM.

No records have been provided relating to internal or external audit of the documentation and data generated during the study undertaken at the test laboratories. The study documentation available to the Rapporteurs does not clearly set out the quality systems used by the manufacturer and/or statisticians to verify, collate and quality assure the data supplied by the three laboratories.

4. Quality of data used for the purpose of the study (existing and newly generated)

4.1 Overall quality of the evaluated testing data (newly generated or existing)

The Rapporteurs believe that although the laboratories which undertook the testing attempted to follow the SOP to the best of their abilities, there were initial proficiency issues; problems with compliance with the SOPs; and problems with the way re-testing was finessed in the analysis, calculations and conclusions included in the 2015 manufacturer's Report to EURL ECVAM.

Supplementary information obtained from the test developer in September 2014 sets out a number of potential deviations from the SOPs known or suspected to have taken place at the test laboratories when the data was generated for the 2011 submission. These include:

- test-kit transport issues;
- differences in the way the test chemicals were stored, handled, and applied to the air-tissue interface;
- issues with the way the chemicals were placed on the test-plates; and
- differences and difficulties with the procedures used to fully remove the test material from the air-tissue interface after the required exposure time.

These insights were not fully disclosed in the 2013 submission to EURL ECVAM, but were used by the test developer to trouble-shoot possible proficiency issues impacting on the reproducibility of the test method, and reinforce to the test laboratories how these crucial protocol steps must be performed. The manufacturer did not consider that the SOP needed to be revised – merely that the laboratories follow best practice and fully comply with the provisions of the SOP.

In addition, the decision that two of the test laboratories would only re-test the more challenging test materials has made it more difficult to determine the likely performance of the test method in fully trained and competent laboratories.

Mention is made elsewhere in this report on the selection of the chemical set for supplementary training and proficiency testing at the two laboratories generating additional data for the 2015 Report to EURL ECVAM.

Whilst there is arguably a case for judgements of test method performance to exclude non-qualified test runs which are known to have been due to technical problems, the PS currently make no provision for this. This is something which may merit consideration when the PS are next revised.

4.2 Quality of the reference data for evaluating reliability and relevance⁶

The relevant PS (EURL ECVAM, 2009; OECD, 2015a) contain provisions to the effect that the chemicals used to develop and optimise a test method should not then be used in a validation study to assess the reproducibility or accuracy of the test method. In this and other cases, contrary to the provisions of the PS, the majority of the 20 reference chemicals were used for test method development/optimisation and for validation. This is not discussed in the 2011 or 2013 submissions to EURL ECVAM. However, the Rapporteurs note that for other RhE "me-too" test methods of the

⁶ OECD guidance document No. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

same class which are now included in TG 439 this provision has not been satisfied either. This is something that might merit consideration when the PS are next revised.

Although the studies which generated the 2011 and 2013 Test Submissions, and the 2015 manufacturer's Report to EURL ECVAM, were reported to have been undertaken in compliance with the agreed protocol, the findings support the conclusion that at the time the data for the 2011 Test Submission was generated, the laboratories, and in particular the naïve laboratory, were not fully proficient. This tends to be confirmed (see above) by supplementary information disclosed by the test method developer in September 2014: there were substantial differences and difficulties with a number of crucial protocol steps when the study to generate the data used in the 2011 submission was undertaken, primarily, but not exclusively, at the naïve laboratory. The test developer used this information to identify key differences and difficulties, and sought to ensure these were remedied before the supplementary testing was performed.

After remedial action had been taken to improve proficiency and SOP compliance, supplementary testing of the full set of PS RCs (n=20) at the naïve laboratory improved the reported reproducibility which met the target value as outlined in the PS.

On that basis the Rapporteurs are satisfied that a plausible partial explanation for the poor reproducibility in the 2011 ring trial was the poor proficiency.

The Rapporteurs also considered whether the 2011 and 2013 test data provided grounds for it being plausible that the poor reproducibility reported in 2011 was instead the result of high intrinsic variability of the test-kits. However, had that been the case the measures to improve proficiency should not of themselves have changed the re-testing results and the problems with reproducibility would not have been restricted to the same subset of test chemicals in all three laboratories.

The Rapporteurs also have concerns about the other elements of the design of the supplementary testing which was performed following the poor reproducibility of the test method report in the 2011 ring trial, and the way the datasets were used to derive the reliability of the performance figures submitted in support of the Test Submission received by EURL ECVAM from the test developer in 2013.

In summary, in 2013 the naïve laboratory, after formal training and proficiency testing, re-tested the original 20 RCs. The other two laboratories re-tested the six chemicals responsible for most of the reproducibility errors observed, and a seventh chemical in one of the laboratories only because it generated invalid run sequences during the 2011 ring-trial. The results of the supplementary testing undertaken by all three laboratories were then substituted for the results obtained for the same chemicals in the same laboratories as reported in the 2011 submission to EURL ECVAM, and it is on that basis revised Reproducibility and Accuracy figures were calculated and supplied to EURL ECVAM by the test developer in 2013.

The selective re-testing and the substitution of data seem to violate the PS' provisions with respect to selective re-testing and data selection. As such in 2014 the Rapporteurs believed the data generated in the second round of testing should be used to support the 2011 submission in view of considering whether the problems encountered had been convincingly resolved on the basis of the hypotheses offered and the measures taken. However the data should **not** be used to selectively *replace* previous data.

The Rapporteurs accept the 2013 Test Submission data as tending to confirm that the origin of the reproducibility problems in the 2011 submission was the imperfect proficiency and competency of the laboratories, and that it also provides useful insights into need for training and proficiency testing being required to ensure laboratories can competently perform the test.

However, the Rapporteurs were not convinced that the reported Reproducibility of the epiCS® test method as set out in the 2013 Test Submission constitute the most relevant, representative or reliable dataset to use to interpret the findings of this study.

The Rapporteurs believe two features of the 2013 study design represent potential sources of optimism-bias, each of which has the potential to inflate the calculated Reproducibility and Accuracy values.

- First: the PS state that for validation studies, chemicals used to develop and optimise the test method should not then be used for the validation study. In this case the 20 RCs used for the study are those recommended in the PS (EURL ECVAM, 2009; OECD, 2015a), and the majority of these chemicals had also been used to develop and optimise the epiCS® test method.
 - However, the Rapporteurs are aware that this provision was not observed by other validation studies of “me-too” methods of the same class that have been validated and accepted for regulatory use.
- Second: with respect to the supplementary testing conducted by the two “non-naïve” laboratories, the Rapporteurs feel that this was not the best approach: interpreting the 2011 and 2013 data sets would have been considerably easier had the full set of chemicals been re-tested by all laboratories and not only by the naïve laboratory.

In 2014 the Rapporteurs advised that at that time a more appropriate, but still imperfect, dataset on which to present and judge to the likely performance of the epiCS® method using the data available from this validation study is to combine and analyse the 2013 data from only the naïve laboratory (as we believe this element of the supplementary testing did address the most plausible avoidable errors introduced by the initial failure to incorporate formal training and proficiency testing and was justified) with the original 2011 datasets from the other two laboratories. On the basis of this re-calculation by the Rapporteurs the only values which fall short of the minimum acceptable value (90%) set out in the PS are the WLR values produced for two of the three laboratories (84% and 84%).

Following this approach the performance parameters were as follows (see Table 2, above):

- The re-calculated WLR values produced on this basis were 95%, 84% and 84%. The minimum acceptable value in the Performance Standards is 90%.
- However, when including also the supplementary data (2013) from the non-naïve laboratories the values were 95%, 100% and 95%, tending to confirm that with experience and full compliance with the SOP acceptable WLR values can be obtained.

The problems with the data analysis in the 2015 Report by the test developer to EURL ECVAM are discussed elsewhere.

- The re-calculated Between Laboratory Reproducibility (BLR) value was 88% (with consistent predictions for 15 chemicals out of 17 chemicals with valid predictions in all three laboratories). The minimum acceptable value in the PS is 80%. The equivalent figure in the 2013 Test Submission was 95% (see Table 3, above).
- With respect to accuracy, the re-calculated values produced were: Sensitivity 95%, Specificity 79%, and Overall Accuracy 87%. These values satisfy the acceptance values set out in the Performance Standards (80%, 70%, and 75%). It is noted that already the 2011 study met the acceptance criteria for predictive capacity.

The Rapporteurs believe that ESAC Opinions and EURL ECVAM decisions must be based on what can be inferred about the values that may have been achieved had the original study design not been compromised by a failure to implement and transfer/training/proficiency phase, and had the full set of test chemicals been re-tested by all three laboratories after the potential proficiency issues been resolved.

One of the relevant PS makes provision for minor deviations from the acceptance criteria⁷ - the other does not. In this case it is plausible both that the initial failure to include a transfer, training and proficiency phase resulted in a lack of proficiency in performing crucial elements of the SOPs at the test laboratories in 2011; and that the 2013 test data, generated after remedial action had been taken, tends to support the hypothesis that subject to attaining proficiency appropriate reproducibility can be attained.

On that basis the Rapporteurs accept that it is plausible the minimum Performance Standards acceptable values for WLR, BLR, Specificity, Sensitivity and Overall Accuracy can be met by trained and competent laboratories – and that formal training and proficiency testing should be required.

The 2015 manufacturer's Report to EURL ECVAM supplied additional test data in line with the 2014 ESAC Opinion. The Rapporteurs have concerns about the quality of the data and analysis set out in the 2015 manufacturer's report to EURL ECVAM: at IIVS more than 5 test runs were required to produce a completed test run sequence for two of the chemicals tested. However, the PS allow a maximum of 5 test runs. Only 5 test runs for each of these chemicals were tabulated in the EURL ECVAM reporting template claiming that three complete test run sequences had been obtained – the test method developer claimed these to be completed test run sequences, and produced test method performance data based on that assumption.

The additional non-qualified test runs ("trials" 2 and 5 at IIVS) were not reported on the EURL ECVAM template, and were not taken into account when the performance figures were calculated by the test developer. One or both of these two additional test runs may have failed due to technical problems (in one case the failure was due to a PC value outside the required range, and in another bacterial/fungal contamination and also a PC value outside the acceptable range). These test runs seem to have been disregarded when the performance figures were calculated, having been considered to be "non-valid" rather than "non-qualified" test runs. The current PS make no provision for making this distinction or treating the results in this way.

For that reason for their evaluation of test method performance the Rapporteurs commissioned a supplementary analysis of the data using a data-set adjusted to comply with the PS re-testing provisions.

4.3 Sufficiency of the evaluated data in view of the study objective

This is discussed fully at 4.2 above.

⁷ For example, the 2009 ECVAM Performance Standard (page 5) states "...it is conceivable that deviations from these standards may be justified by scientific reasons...". The OECD Performance Standard does not contain this provision.

5. Test definition (Module 1)

5.1 Quality and completeness of the overall test definition

All of these are clearly set out in the 2013 Test Submission, and in 2009 was deemed by EURL ECVAM (then ECVAM) to satisfy the requirements of the essential test method components of the relevant Performance Standards.

Between the initial ring-trial and the supplementary testing, the Epidermal Skin Test EST-1000, was renamed "epiCS® SIT". EURL ECVAM requested and received a formal assurance from the submitter that only the name but no other parameter of the test method had been subject to change.

The Rapporteurs noted that during the validation study the nature of the plastic inserts which form part of the test kit were changed, but that this did not compromise test performance values. The Rapporteurs consider that the ability of the test method to withstand such a change in manufacturing process tends to confirm the robustness of the test method and its external validity.

5.2 Quality and completeness of the documentation concerning SOPs and prediction models

The Rapporteurs are satisfied that the 2013 Test Submission, and supporting reference documents including the supplementary information supplied by the test developer in September 2014, are sufficiently detailed and complete on these points. The definitive SOP is "INVITTOX Protocol epiCS® SKIN IRRITATION TEST Standard Operating Procedure (SOP epiCS SIT) Version 4.0 November 2012".

The SOP contains a sub-protocol (Interference Test) invoked for chemicals (strongly coloured and/or MTT reducers) which may not otherwise be compatible with or properly categorised by the epiCS® test method. The reproducibility and reliability of this sub-protocol were not evaluated during this validation study.

6. Test materials

6.1 Sufficiency of the number of evaluated test items in view of the study objective

The 20 RCs used are the minimum base-set as specified in the PS (EURL ECVAM, 2009; OECD, 2015a). As such they are sufficient in number and suitably representative and balanced with respect to different chemical classes and physical states, and skin irritation potential as determined by the best available reference test methods and reference data.

The majority of these chemicals has also been used to develop and optimise the epiCS® test method. In these circumstances within the PS framework provision is made for their substitution by chemicals not used for test method development or optimisation. This option, had it been exercised, would have better met the letter and spirit of the Performance Standards.

6.2 Representativeness of the test items with respect to applicability

The 20 RCs used are as listed in the relevant PS, and are suitably representative and balanced with respect to different chemical classes and physical states, and skin irritation potential as determined by the best available reference test methods and data.

7. Within-laboratory reproducibility (WLR) (Module 2)

7.1 Assessment of repeatability and reproducibility in the same laboratory

See Section 4.2 – specifically the final dataset relied upon by the Rapporteurs.

7.2 Conclusion on within-laboratory reproducibility as assessed by the study

See Section 4.2 – specifically the final dataset relied upon by the Rapporteurs, which shows a re-calculated value (88.89%) for one laboratory that falls fractionally short of the PS criterion of $\geq 90\%$.

8. Transferability (Module 3)

8.1 Quality of design and analysis of the transfer phase

The study which generated the data in support of the 2011 Test Submission did not include a formal transferability/training/proficiency testing phase. In the view of the Rapporteurs this was a significant omission, and is likely to be the main cause of the 2011 study's failure to meet the required reproducibility targets.

Remedial action was undertaken prior commencing the supplementary round of re-testing, but full details of what was involved were only disclosed by the test method developer in September 2014.

8.2 Conclusion on transferability to a naïve laboratory / naïve laboratories as assessed by the study

The study which generated the data in support of the 2011 Test Submission did not include a formal transferability/training/proficiency testing phase. The WLR for the naïve laboratory for that round of the validation study was 74%, the values for the non-naïve laboratories were both 84%.

After remedial action was taken in an attempt to improve proficiency the naïve laboratory re-tested the full set of 20 RCs the WLR increased to 95%. This tends to confirm that, subject to suitable transfer/training/proficiency testing, the epiCS® method can be successfully transferred to a naïve laboratory – and that this should be required of any new laboratory seeking to use the test method.

The 2011 WLR at the two non-naïve laboratories (84% in each laboratory) was below the minimum acceptable PS value of 90%. Values of 95% and 100% were obtained when the 2013 re-testing data was substituted for the 2011 data: this suggests that even laboratories experienced and proficient in the use of similar test methods require training and proficiency testing if planning to perform the epiCS® method.

In addressing the 2014 ESAC recommendation regarding additional testing, the two non-naïve laboratories underwent additional training and proficiency testing using 5 test chemicals, all of which were then to be re-tested to generate the required test data for the 2015 test developer Report to EURL ECVAM. This is not best practice: none of the chemicals to be tested should have been used for this supplementary round of proficiency testing.

9. Between-laboratory reproducibility (BLR) (Module 4)

9.1 Assessment of reproducibility in different laboratories

See Section 4.2 – specifically the final dataset relied upon by the Rapporteurs.

9.2 Conclusion on between-laboratory reproducibility as assessed by the study

See Section 4.2 – specifically the final dataset relied upon by the Rapporteurs which estimates BLR at 88.24% (with a PS target of $\geq 80\%$).

10. Predictive capacity and overall relevance (Module 5)

10.1 Adequacy of the assessment of the predictive capacity in view of the purpose

The predictive capacity values were calculated using the computational model provided in the relevant PS.

The predictive capacity figures provided in both the 2011 and the 2013 Test Submissions appeared to satisfy the minimum acceptable values for test method Sensitivity, Specificity, and Overall Accuracy.

However, the Rapporteurs do not believe that the reported reproducibility and accuracy of the epiCS® test method as set out in either the 2011 or 2013 Test Submissions are the most relevant and reliable datasets to use to interpret the findings of this study with respect to reliability, and have concerns about the design of the supplementary testing which was performed following the poor reproducibility of the test method report in the 2011 ring trial, and the way the data was then substituted for the original ring-trial data. The details of the supplementary testing and data analysis, and possible sources of bias, are described above.

In 2014 the Rapporteurs reasoned that a more relevant and appropriate dataset on which to present and judge to the performance of the epiCS® method using the data available from this validation study is to combine and analyse the 2013 data from the naïve laboratory (as we believe this element of the supplementary testing did address avoidable errors introduced by the initial failure to incorporate formal training and proficiency testing and was justified) with the original 2011 datasets from the other two laboratories.

When that is done:

TABLE 6: Predictive Capacity as described in the two submissions (2011 and 2013) and as re-calculated by the ESAC Rapporteurs and EURL ECVAM.

	2011	2013	2014 ESAC WG re-derived value	Performance Standard Minimum Acceptable Value
Sensitivity	93%	97%	100%*	80%
Specificity	78%	80%	79%	70%
Overall Accuracy	85%	88%	89%	75%

**) the number is higher than the 2013 values since in 2013 HCCR had reported one FN, which reduces the Sensitivity. When not considering the 2013 re-testing data from IIVS and HCCR, but combining the 2011 data from these two labs with the 2013 re-testing data from ACS (all 20 chemicals), there is no FN in the data matrix.*

On that basis the Rapporteurs were satisfied that for all reasonable Predictive Capacity values that can be derived meet the minimum acceptable value set out in the PS.

Based on the 2013 test data only one Category 2 material (1-methyl-3-phenylpiperazine) was misclassified as a non-irritant by one laboratory. All three laboratories derived appropriate classifications for the Category 2 substances 1-decanol and di-n-propyl disulphide. All laboratories misclassified the non-irritants allyl phenoxy acetate and cinnamaldehyde as irritants.

In a letter of September 2014, the test developer speculated, but provided no evidence, to test material oxidation being a possible confounding factor.

In its 2015 Report to EURL ECVAM in response to the 2014 ESAC Opinion consideration of supplementary data and analysis purported to show the PS requirements for Relevance and Accuracy has been met: the same is true when the judgement is made on the basis of the revised calculation commissioned by the Rapporteurs (see 4.2)

TABLE 7: Predictive Capacity as described in the 2015 manufacturer’s Report to EURL ECVAM, and the further analysis commissioned by the Rapporteurs.

	2015 Manufacturer’s Report to EURL ECVAM	2015 Rapporteurs’ re- worked analysis	Performance Standard Minimum Acceptable Value
Sensitivity	86.21%	88.89%	80%
Specificity	80.00%	80.00%	70%
Overall Accuracy	83.05%	84.21%	75%

10.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

The Rapporteurs consider that the 2013 Test Submission and the supporting documents confirm the biological/mechanistic relevance, and relevance in terms of making accurate measurements and predictions for the skin irritation potential endpoints set out in the study objectives: and that the general assertions made to this effect in the 2013 Test Submission are substantiated by the data generated during the validation study and adjunct testing.

The Rapporteurs also consider the performance criteria set out in the 2015 manufacturer’s report to EURL ECVAM did not give proper consideration to the reporting and analysis of non-qualified test runs, and rely instead on a supplementary analysis commissioned to address this problem.

11. Applicability domain (Module 6)

11.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

On the basis of the available evidence (as at May 2016) the Rapporteurs conclude that the applicability domain of the epiCS® test method can be expected to match that of the validated RhE *in vitro* test methods from which the relevant PS were derived.

The chemicals which were inappropriately classified in course of the study tend to be those which have been problematical with the other validated RhE *in vitro* test methods.

11.2 Quality of the description of applicability domain, limitations, exclusions

These are only briefly discussed in the 2013 Test Submission. See 11.1.

12. Performance standards (Module 7)

This section is not applicable to the study.

12.1 Adequacy of the proposed Essential Test Method Components

Not applicable – see immediately above.

12.2 Adequacy of the Reference Chemicals

Not applicable – see immediately above.

12.3 Adequacy of proposed performance target values

Not applicable – see immediately above.

13. Readiness for standardised use

13.1 Assessment of the readiness for regulatory purposes

The transfer/training/proficiency requirements need to be defined, and made pre-requisites for the use of this test method. An SOP for training and transfer should be agreed/developed for this purpose.

The sub-protocol on interference test (to be used with Colour interfering and/or direct MTT reducing test chemicals) has yet to be validated.

13.2 Assessment of the readiness for other uses

The Rapporteurs see no reason in principle why trained and competent laboratories should not use this test method for relevant non-regulatory purposes.

13.3 Critical aspects impacting on standardised use

The Rapporteurs consider training and proficiency testing to be essential requirements, and understand that the manufacturer can and will assist with this. In addition an appropriate SOP for transferability and proficiency testing is required.

13.4 Gap analysis

Training and proficiency testing are essential requirements, and the manufacturer can and will assist with this. In addition an appropriate SOP for transferability and proficiency testing is required.

14. Other considerations

The Rapporteurs are grateful for the information provided by the test developer, notwithstanding the fact that some relevant information was only disclosed on request, and further analysis was required to bring the analysis set out in the 2015 manufacturer's Report to EURL ECVAM into compliance with the PS requirements relating to re-testing.

The quality and completeness of the background material provided, and the presentation of the findings, are generally of a good standard.

There are issues relating to experimental design; and the data reporting, data analysis, and conclusions supplied by the manufacturer. These are covered in detail elsewhere in this report.

Each RC was evaluated in at least three independent runs in each laboratory, with the test runs conducted at different time points and using different tissue batches. The precise requirements for "independence" of runs are not specified in the Performance Standards: some of the "independent" test runs in this validation were separated by short time intervals, and some of the test-kits were from consecutive of very close test-kit production runs. The meaning of "independence" might be clarified in future PS.

Arguably in determining test method performance test runs which are currently classed as "non-qualifying" due to technical reasons (for example PC values outside the required range, or bacterial/fungal contamination) might reasonably be set aside as "not valid" rather than included as "non-qualifying runs" and consequently originating non-qualified test for all test chemicals included in those runs; however, to avoid data-selection this should only be considered if the technical issues are identified and the test run aborted before test data on the RCs is collected. At present the PS make no provision for this. This might merit consideration when the PS are next revised.

There are occasions when failure to produce a complete test run sequence (three qualified test runs) for more than one test chemical in one laboratory leads to those chemicals being excluded from the WLR and BLR calculations. There may be a case for considering whether in cases where this markedly reduces the number of chemicals included in the final analysis, mindful that these may be the more challenging test chemicals, that the requirement for three qualified test runs in each laboratory might be reduced to one or two for assessing BLR. Before considering whether this should be considered when the PS are next revised, some existing datasets might be reworked to determine what effect this might have on the calculation of test method performance indices.

15. Conclusions on the study

15.1 ESAC WG summary of the results and conclusions of the study

See Executive Summary.

15.2 Extent to which study conclusions are justified by the study results alone

See Executive Summary.

15.3 Extent to which conclusions are plausible in the context of existing information

See Executive Summary and Section 4.2.

16. Recommendations

16.1 General recommendations

The Rapporteurs offer the following draft recommendations for consideration:

1. The epiCS[®] RhE test method can only be deemed to satisfy the minimum performance criteria (for Reproducibility and Predictive Capacity) requirements set out in the EURL ECVAM/OECD Performance Standards for *in vitro* skin irritation testing using *Reconstructed human Epidermis (RhE)* (EURL ECVAM, 2009; OECD, 2015a), to discriminate between “non-irritant” (No Category = not-classified) and classified chemicals ("category 1" or "category 2") for skin corrosion/irritation, and be “sufficiently similar” to the existing Validated Reference Methods, if account is taken of the PS provision that minor deviations (in this case a minor shortfall in the estimated WLR value) may be accepted for scientific reasons, and the poor reproducibility values reported for the 2011 study were due to poor proficiency on the part of the untrained laboratories conducting the tests.
2. The OECD TG439 (2013 and 2015) has no similar provision, and consideration should be given to its being updated to take account of this.
3. The use of the epiCS[®] RhE test method for regulatory purposes for skin irritation testing be conditional upon all laboratories demonstrating that they have undertaken appropriate training and proficiency testing. The test method SOP should be amended to make provision for this. This might be considered for other “me-too” RhE test methods.
4. Recommended to emphasize in SOP:
 - a. importance of washing steps;
 - b. keeping chemicals separated during exposure and post-incubation
5. The Rapporteurs believe that any revised PS should contain more explicit provisions and advice for how studies should be designed and conducted if any of the chemicals listed as Performance Standards Reference Chemicals were also used to develop or optimise the test method.

16.2 Specific recommendations (e.g. concerning improvement of SOPs)

A transfer, training and proficiency SOP is required.

17. References

Draize J.H., Woodand G. & Calvery H.O. (1944). Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *Journal of Pharmacology and Experimental Therapeutics*. 82, pp 337-390.

EURL ECVAM (2009). Performance Standards For In-Vitro Skin Irritation Test Methods Based On Reconstructed Human Epidermis. Online:
http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/validation-regulatory-acceptance/docs-skin-irritation-1/DOC8-updated_ECVAM_PS_2009.pdf

Griesinger, C., Barroso, J., Zuang, V., Cole, T., Genschow, E. and Liebsch M. (2009). Explanatory background document to the OECD draft test guideline on in vitro skin irritation.
Online:
http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/validation-regulatory-acceptance/topical-toxicity/skin-irritation
http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/validation-regulatory-acceptance/docs-skin-irritation-1/DOC10-background%20document.pdf

OECD (2002). OECD Guideline for Testing of Chemicals, No. 404: Acute Dermal Irritation/Corrosion. Organisation for Economic Cooperation and Development, Paris.

OECD (2005). OECD Guidance Document 34 — Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment.

OECD (2010). OECD Guideline for Testing of Chemicals, No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method.

OECD (2013). OECD Guideline for Testing of Chemicals, No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method.

OECD (2015a). Performance Standards for the assessment of proposed similar or modified *in vitro* Reconstructed human Epidermis (RhE) test methods for skin irritation testing as described in TG 439 (Intended for the developers of new or modified similar test methods). Series on Testing and Assessment No. 220.

OECD (2015b). OECD Guideline for Testing of Chemicals, No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method.

Tornier, C., Roquet, M., Fraissinette, A.B. (2010). Adaptation of the validated SkinEthic™ Reconstructed Human Epidermis (RHE) skin corrosion test method to 0.5 cm² tissue sample. *Toxicol. In vitro* 24, 1379-1385.

UN (2015). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Sixth revised edition, UN New York and Geneva.

Europe Direct is a service to help you find answers to your questions about the European Union

Free phone number (*): 00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.

It can be accessed through the Europa server <http://europa.eu>

How to obtain EU publications

Our publications are available from EU Bookshop (<http://bookshop.europa.eu>), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents.

You can obtain their contact details by sending a fax to (352) 29 29-42758.

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub

ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



Joint Research Centre



EU Science Hub